Draft version February 16, 2023
Typeset using LaTeX **twocolumn** style in AASTeX631

arXiv:2302.07277v1 [astro-ph.CO] 14 Feb 2023

On the nature of disks at high redshift seen by JWST/CEERS with contrastive learning and cosmological simulations

Jesús Vega-Ferrero,[1,2,3] Marc Huertas-Company,[1,2,4] Luca Costantin,[5] Pablo G. Pérez-González,[5] Regina Sarmiento,[1,2] Jeyhan S. Kartaltepe,[6] Annalisa Pillepich,[7] Micaela B. Bagley,[8] Steven L. Finkelstein,[8] Elizabeth J. McGrath,[9] Johan H. Knapen,[1,2] Pablo Arrabal Haro,[10] Eric F. Bell,[11] Fernando Buitrago,[3,12] Antonello Calabrò,[13] Avishai Dekel,[14,15] Mark Dickinson,[10] Helena Domínguez Sánchez,[16] David Elbaz,[17] Henry C. Ferguson,[18] Mauro Giavalisco,[19] Benne W. Holwerda,[20] Dale D. Kocesvski,[9] Anton M. Koekemoer,[18] Viraj Pandya,[21,*] Casey Papovich,[22,23] Nor Pirzkal,[24] Joel Primack,[15] and L. Y. Aaron Yung[25]

ABSTRACT

Visual inspections of the first optical rest-frame images from JWST have indicated a surprisingly high fraction of disk galaxies at high redshifts. Here we alternatively apply self-supervised machine learning to explore the morphological diversity at $z \geq 3$. Our proposed data-driven representation scheme of galaxy morphologies, calibrated on mock images from the TNG50 simulation, is shown to be robust to noise and to correlate well with physical properties of the simulated galaxies, including their 3D structure. We apply the method simultaneously to F200W and F356W galaxy images of a mass-complete sample ($M_*/M_\odot > 10^9$) at $z \geq 3$ from the first JWST/NIRCam CEERS data release. We find that the simulated and observed galaxies do not populate the same manifold in the representation space from contrastive learning, partly because the observed galaxies tend to be more compact and more elongated than the simulated galaxies. We also find that about half the galaxies that were visually classified as disks based on their elongated images actually populate a similar region of the representation space than spheroids, which according to the TNG50 simulation is occupied by objects with low stellar specific angular momentum and non-oblate structure. This suggests that the disk fraction at $z > 3$ as evaluated by visual classification may be severely overestimated by misclassifying compact, elongated galaxies as disks. Deeper imaging and/or spectroscopic follow-ups as well as comparisons with other simulations will help to unambiguously determine the true nature of these galaxies.

Keywords: Galaxy formation (595); Galaxy evolution (594); High-redshift galaxies (734); Neural networks (1933);

1. INTRODUCTION

Understanding how galaxy diversity emerges across cosmic time is one of the major goals of galaxy formation. How and when do stellar disks form? What are the main drivers of bulge growth? How and when did galaxy morphology and star formation get connected? Despite significant progress in the past years, thanks in particular to deep surveys undertaken with the Hubble Space Telescope (e.g., Scoville et al. 2007; Grogin et al. 2011a; Koekemoer et al. 2011), these questions remain largely unanswered. The general picture is that massive star-forming galaxies in the past were more irregular in their stellar structure (e.g., Abraham et al. 1996; Conselice 2003) than today's disks even if observed in the optical rest-frame (Buitrago et al. 2013; Huertas-Company et al. 2015). Galaxies above $z \sim 1$ also show the presence of giant star-forming clumps (e.g, Guo et al. 2015, 2018; Huertas-Company et al. 2020; Ginzburg et al. 2021) which might indicate a turbulent and unstable ISM (e.g., Ceverino et al. 2010; Bournaud et al. 2014). Although the gas shows signatures of rotation at $z \sim 2$ (e.g.,Wisnioski et al. 2015), the settling of disks seems to be a process happening at least from $z \sim 2$ (e.g., Kassin et al. 2012; Buitrago et al. 2014; Simons et al. 2017; Costantin et al. 2022a) coincident with the decrease of gas fractions in massive galaxies (e.g., Freundlich et al.

Corresponding author: Jesús Vega-Ferrero
astrovega@gmail.com

* Hubble Fellow

2019; Genzel et al. 2010). Another important result of the past years is that the presence of bulges in galaxies is strongly correlated with the star formation activity at all redshifts probed (e.g., van der Wel et al. 2014a; Barro et al. 2017; Costantin et al. 2020, 2021; Dimauro et al. 2022). This suggests that bulge formation and quenching are tightly connected physical processes (e.g., Chen et al. 2020b).

With its unprecedented sensitivity, spatial resolution and infrared coverage, the JWST is offering a new window to the stellar structure of galaxies in the first epochs of cosmic history (Gardner et al. 2006). For the first time, we are able to explore the stellar morphologies of the first galaxies formed in the universe, which should enable new constraints on the physical processes governing galaxy assembly at early times and hopefully a better understanding of the physical processes leading to the formation of the first stellar disks and bulges. Some very recent works have already started this exploration by performing visual classifications (Ferreira et al. 2022a,b; Kartaltepe et al. 2022) or by applying supervised Machine Learning trained on HST images (Robertson et al. 2023) of galaxies observed in the first JWST deep fields. One of the main results of these early works is that JWST seems to be detecting star-forming disks even at $z > 3$, which would push the time of disk formation to very early epochs. Two questions naturally arise from these first works:

- Are the galaxies seen by JWST true disks, i.e flat, rotating systems? The aforementioned results are based primarily on qualitative morphological classifications, with quantitative tracers of morphology (e.g., Sersic fits) incorporated to further inform differences between the visually defined classes. However, galaxies might look morphologically disky but have significantly different stellar kinematics than local disk galaxies. Distinguishing edge-on flat disks from more prolate systems is also a very challenging task that could bias the results (e.g. van der Wel et al. 2014b; Zhang et al. 2019).

- Do modern cosmological simulations reproduce the observed galaxy diversity at $z > 3$? Although some preliminary comparisons exist, a fair comparison in the observational plane is required to fully address this question (e.g., Rodriguez-Gomez et al. 2019; Huertas-Company et al. 2019; Zanisi et al. 2021).

In this work, we attempt to provide new insights into these two main questions. To that purpose, we apply a novel data-driven approach based on contrastive learning (Hayat et al. 2021; Sarmiento et al. 2021) to a mass complete sample of galaxies at $z \geq 3$ in the CEERS survey (Finkelstein et al. 2022a). By calibrating the method with mock galaxies (Costantin et al. 2022b) from the TNG50 cosmological simulation (Nelson et al. 2019a; Pillepich et al. 2019; Nelson et al. 2019b) and by choosing the proper augmentations (i.e., transformations applied to the images such as rotations, flux normalizations, noise, etc.), we are able to build a morphological description which is more robust to noise and galaxy orientation than more traditional approaches. Our morphological representation can then be correlated with the physical properties of galaxies from the simulation to provide new insights about the physical nature of visually classified disks and to explore the agreements and disagreements between observations and simulations.

The paper proceeds as follows: in section 2 we describe the galaxy datasets used in this work; section 3 describes the contrastive learning setting used to derive unsupervised representations of galaxy morphologies; section 4 explores the properties of the obtained representations on observed JWST/CEERS galaxies; the results and implications are discussed in section 5; finally, a summary and the final conclusions are presented in section 6.

2. DATA

2.1. *CEERS*

We use JWST imaging data from NIRCam obtained during the first epoch (June 21-22, 2022) of the Cosmic Evolution Early Release Science (CEERS; Finkelstein et al. 2017) survey. This consists of short and long-wavelength images in both NIRCam A and B modules, taken over four pointings, labeled NIRCam1, NIRCam2, NIRCam3, and NIRCam6. Each pointing was observed with seven filters: F115W, F150W, and F200W on the short-wavelength side, and F277W, F356W, F410M, and F444W on the long-wavelength side. Here we only use the F200W and F356W filters. These two filters probe the UV, optical and near-IR rest-frame at $z > 3$, allowing to probe simultaneously the distribution of young and old stars. A full description of this release and the data reduction can be found in Bagley et al. (2022) and Finkelstein et al. (2022b).

In addition to the images, we use two different catalogs with physical properties of galaxies:

- CEERS catalog (CEERS): a photometric catalog (Bagley et al. 2022; Finkelstein et al. 2022b) with derived stellar masses and photometric redshifts ($z_{\rm phot}$) obtained through SED fitting of the latest data reduction photometry (Pablo G. Pérez-

González private communication). For a fair comparison with the simulated TNG50 dataset — see subsection 2.2— we select 891 galaxies with $3 \leq z \leq 6$ and stellar masses $M_* \geq 10^9 M_\odot$. We also impose the selected galaxies to have Kron-based fluxes in F200W and F356W filters above zero to avoid spurious sources. Following Pozzetti et al. (2010), we derived a completeness stellar mass of $\sim 10^7 M_\odot$ for the considered redshift range. The subsample of the CEERS catalog used in this work is, therefore, mass-complete within $3 \leq z \leq 6$.

- Visual classification catalog (VISUAL): a redshift-selected $z \geq 3$ morphological catalog presented in Kartaltepe et al. (2022) containing galaxies in common between CANDELS (Grogin et al. 2011b) and CEERS observations. This is intended to directly compare our morphological description to the visual classification of Kartaltepe et al. (2022). Redshifts and stellar masses are extracted from CANDELS v2 for the HST F160W-selected galaxies in the EGS field (see Kodra et al. 2022, for full details on the photometric redshift measurements and resulting catalogs). The morphological catalog presented in Kartaltepe et al. (2022) consists of a redshift-selected sample with 850 galaxies at $z \geq 3$. The visual classifications of each galaxy are performed by three people. A given classification is assigned if two out of three people select that option. Galaxies classified in this way are broken down into the following non-exclusive morphological groups: Disk only, Disk+Spheroid, Disk+Irregular, Disk+Spheroid+Irregular, Spheroid only, Spheroid+Irregular and Irregular only. See Kartaltepe et al. (2022) for more details on the different classification tasks and morphological groups.

Both catalogs also include morphological measurements of Sérsic index (n_e), semi-major axis (a) and axis-ratio (b/a) derived with `galfit` (Peng et al. 2010). More information about the fits can also be found in Kartaltepe et al. (2022).

The distributions of stellar masses of the galaxies in the CEERS and the VISUAL datasets are shown in Figure 1.

2.2. *Mock JWST images of TNG50 galaxies*

We use the TNG50-1[1] suite of simulation (hereafter TNG50; Nelson et al. 2019a; Pillepich et al. 2019; Nelson et al. 2019b) and their mock NIRCam observations at $z > 3$ galaxies following the observational strategy of CEERS. The mock images[2] were produced by modeling the gas cells and star particles in the simulation as presented in Costantin et al. (2022b). We consider four snapshots of the TNG50 simulation corresponding to $z = (3, 4, 5, 6)$ and galaxies with stellar masses $M_* \geq 10^9 M_\odot$. In total, the original dataset consists of 1326 galaxies (see Table 1). Each selected galaxy is then observed along 20 different line-of-sight orientations to increase the statistics which produces a dataset of 26520 galaxy images that we consider as independent objects for the purpose of this work. As described in Costantin et al. (2022b), parametric and non-parametric morpho-

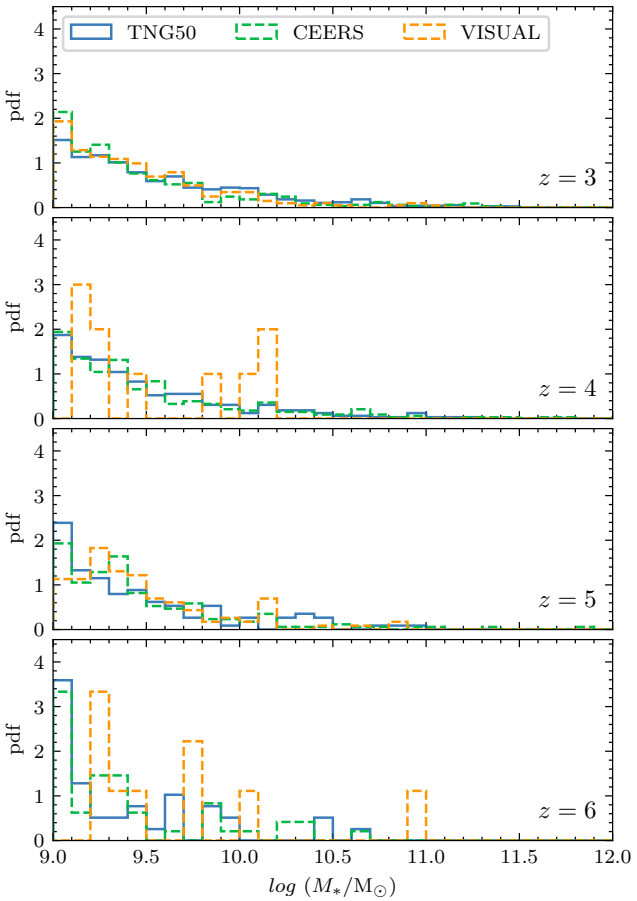

Figure 1. Probability density function of the logarithm of the stellar mass of the simulated galaxies in the TNG50 dataset (blue histogram), and the observed galaxies in the CEERS and VISUAL datasets (green and orange dashed histogram, respectively). Different panels correspond to the redshifts analyzed, $z = (3, 4, 5, 6)$.

[1] https://www.tng-project.org/

[2] Data publicly released at https://www.tng-project.org/costantin22

logical parameters for this dataset are derived using the standard configuration of `statmorph`[3](Rodriguez-Gomez et al. 2019).

For this work, we use the noiseless images in the F200W and F356W bands from the Costantin et al. (2022b) dataset with a pixel scale of 0.031 and 0.063 arcsec pix^{-1}, respectively. The field-of-view of each image is equal to the total (dark matter, gas and star particles included) half-mass radius of the corresponding galaxy. Our image classification scheme requires a fixed image size. Therefore, we select galaxy images with a field-of-view larger than 64×64 and 32×32 pixels in the F200W and F356W bands, respectively, and generate cutouts of those sizes. Then, to match both observations of the same galaxy, images in the F356W band are re-sampled to the same pixel scale as the F200W images. According to these criteria, $\lesssim 7\%$ of the galaxies (most of them at $z = 3$) are dropped out from our initial sample. The total number of galaxies considered is finally 1 238 distributed within $z = 3-6$ (see Table 1), which translates into 24 760 projections. Although the number of objects we remove is small, we check in Figure 2 if a specific population is systematically removed. The figure shows the size-mass relation of the selected TNG50 dataset along with the excluded galaxies based on the size of the field-of-view. The excluded galaxies are not necessarily the most compact and/or less massive galaxies in the dataset. However, a fraction of them with lower-than-average stellar extent are indeed removed based on our selection and would reach otherwise sizes of a few hundreds of parsecs.

For comparison, we show in Figure 1 the distribution of the stellar masses in the simulated TNG50 dataset, and the observed CEERS and VISUAL datasets. Note the good agreement between the TNG50 and the CEERS samples, even if we are comparing here the stellar masses directly extracted from the TNG50 simulations and those obtained through SED fitting of the latest JWST data. Also remarkable is the agreement, despite selection effects, between the TNG50 and the VISUAL datasets.

3. SELF-SUPERVISED LEARNING REPRESENTATION OF MOCK JWST IMAGES OF SIMULATED TNG50 GALAXIES

In this section, we describe the main methodology we develop to obtain a data-driven morphological description of galaxies that is robust to noise and other nuisance parameters.

[3] `statmorph` is available at https://statmorph.readthedocs.io.

Table 1. Summary of the sample of simulated galaxies from the TNG50 dataset. The first column indicates the redshift (z); the second column shows the total number of galaxies in the simulated TNG50 dataset; the third column refers to the number of selected galaxies according to image size limitations (i.e., 64×64 and 32×32 pixels in the F200W and F356W bands, respectively).

z	All galaxies	Selected galaxies
3-6	1326	1238
3	829	760
4	343	326
5	115	113
6	39	39

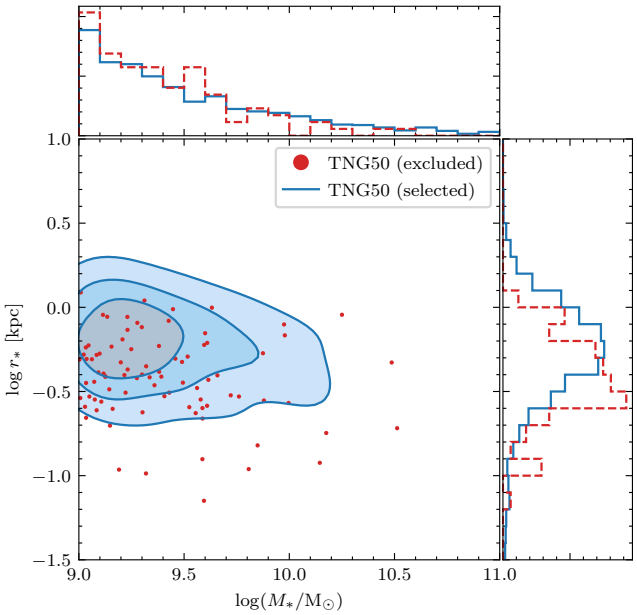

Figure 2. Logarithm of the physical size (stellar half-mass radius, r_*, in kpc) versus the stellar mass (M_*, in M$_\odot$) of the TNG50 galaxies. Blue-filled contours show the 25%, 50% and 75% probabilities of the TNG50 selected galaxies. Red data points correspond to the excluded galaxies in terms of the size of the field-of-view. See Table 1.

3.1. *Contrastive learning framework*

Our approach is based on an adaptation of the Simple framework for Contrastive Learning of visual Representations (SimCLR; Chen et al. 2020a). Very briefly, the idea behind the SimCLR framework is to obtain robust representations of images without labels by applying random augmentations as explained below.

Given an image, random transformations are applied to it to generate a pair of two augmented images, (x_i, x_j). Each image in the pair is passed through a Convolutional Neural Network (CNN) to compress the images into a set of vectors, (h_i, h_j). Then a non-linear

fully connected layer (i.e., projection head) is placed to get the representations (z_i, zj). The representations are learned iteratively by maximizing agreement between the augmented views of the same image example (z_i, zj) and minimizing agreement between all other pairs considered as negative. This is achieved via a so-called contrastive loss in the latent space, :

$$l_{i,j} = -\log \frac{\exp(\langle \boldsymbol{z}_i, \boldsymbol{z}_j \rangle / h)}{\Sigma_{k=1,k\neq i}^{2N} \exp(\langle \boldsymbol{z}_i, \boldsymbol{z}_k \rangle / h)}, \quad (1)$$

where $\langle \boldsymbol{u}, \boldsymbol{v} \rangle$ denotes the dot product between L^2-normalized $\boldsymbol{u}$ and $\boldsymbol{v}$, and h denotes the temperature parameter that regulates the distribution of the output representations (see Hinton et al. 2015; Wu et al. 2018, for more details). The final loss is computed in batches of size N across all positive pairs, both (i, j) and (j, i), while the rest of the augmented examples are treated as negative examples, which are denoted by k.

For this work, we follow the implementation from Sarmiento et al. (2021) which was successfully applied to astronomical data. The CNN encoder consists of four convolutional layers with kernel sizes 5, 3, 3, 3, and 128, 256, 512 and 1 024 filters per layer, respectively. Max-pooling layers and Exponential Linear Unit (ELU) activation functions are placed after each convolutional layer. Therefore, the representations before the projection head —(h_i, h_j)— for each galaxy image are encoded into 1 024 features. Subsequently, the projection head (composed of three fully connected layers of 512, 128 and 64 neurons per layer) transforms the galaxy representations to a latent space —(z_i, z_j)— where the contrastive loss is computed.

3.2. *Data augmentation and network training*

The choice of data augmentations is a key element in contrastive learning training (Chen et al. 2020a) as it allows us to turn the representations independent of some nuisance effects. In the context of this work, our goal is to obtain a morphological representation that is robust to signal-to-noise, rotation, size and does not depend on color. To reach this objective, we calibrate our algorithm on the mock TNG50 dataset, since it allows us to access noiseless versions of the images and, therefore, marginalize over the noise. More precisely, we apply the following augmentations:

- *Noise*: as described in section 2, for each galaxy image in the F200W and F356W bands we have a noiseless version that does not include any instrumental effects or noise. Using the available CEERS data, we construct mock CEERS galaxy images as a combination of the TNG50 noiseless images and random patches of the four observed CEERS pointings. We first add source Poisson noise before convolving the images with a PSF extracted from the observations (Finkelstein et al. 2022b) in each band. Then, we add real-time realistic noise (that may also include other sources/interlopers) to each of the 64×64 noiseless galaxy images by summing up randomly chosen patches from the CEERS pointing of the same size. From the contrastive learning point of view, these images with a real background are considered as an augmented copy of their noiseless analogs during the training process. These augmentations should enforce the representations to be robust to signal-to-noise (S/N) as well as to background and foreground companions.

- *Rotation*: we apply a random flipping (horizontal or vertical, but not both) and a random rotation with 100% chance independently to both the TNG50 noiseless image and the patch of the sky extracted from the CEERS pointings. This augmentation is intended to ensure the model is invariant to the galaxy orientation;

- *Flux*: we randomly apply independent flux scaling to the noise-added images. The associated flux is random but follows the flux distribution of the TNG50 dataset in the F356W band. We randomly sample flux values from the TNG50 F356W flux distribution and apply them accordingly to noise-added images in the two filters. This augmentation is intended to stress the robustness to S/N. It also helps —as we will show in the following— to make the representations independent of galaxy size since the regions above the noise scale will vary with the flux variations. We note that we decided not to apply direct size augmentation, (i.e. such as zoom-in or zoom-out), as that would force us to up-sample or down-sample the images with less or more than 64×64 pixels size, respectively, which creates some artifacts that the network learns.

- *Color*: finally, in order to prevent the network from learning color information and/or the intrinsic brightness of the galaxies directly from the images, we apply two additional augmentations to both the noiseless and the noise-added images. First, each band is normalized individually after the augmentations are applied. Second, the noiseless and noise-added images are normalized independently and individually for each galaxy. Consequently, the maximum pixel value in every galaxy

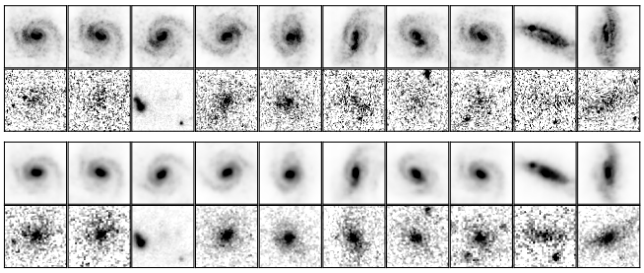

Figure 3. Images of a TNG50 galaxy at $z = 3$ with $M_* \approx 6 \times 10^9 M_\odot$. The first two rows show noiseless and augmented versions, respectively, of 10 different projections of the selected galaxy in the F200W filter. The third and fourth rows correspond to the noiseless and augmented versions, respectively, of the same galaxy projections shown in the first two rows in the F356W filter. All the panels have a 64×64 pixels size. The pixel values have been *asinh* transformed with a 0.5% clipping.

image is equal to one for each band and counterimage.

In Figure 3, we show several projections of a galaxy at $z = 3$ with $M_* \approx 6 \times 10^9 M_\odot$ extracted from the TNG50 simulation. In some of the augmented versions (along with PSF convolution, realistic noise, random rotations, flux variation, etc.) it is possible to distinguish several companions within the stamps. This is the result of adding randomly chosen patches of the CEERS pointing to the TNG50 noiseless images. These examples come from an extended bright galaxy that appears significantly weaker in the augmented images due to the flux variations applied in the augmentations and given that the TNG50 distribution of F356W peaks at lower values (compared to the selected galaxy).

To reduce the dynamic range and to be sensitive to both the center and outskirts of the galaxy, before training, we apply a *asinh* (inverse hyperbolic *sin*) transformation and a minimum-maximum normalization to each galaxy pair in each band. Additionally, we ensure that only one projection of the same galaxy enters each batch during training and that all the galaxy images are passed through the network at every epoch. We do so to avoid the algorithm learning the orientation of the same galaxy as seen from different line-of-sight projections since some of the projections are just a simple rotation of the galaxy in the sky.

Our contrastive SimCLR model is trained with the mock JWST images for 24 760 different projections of 1 238 galaxies within $3 \leq z \leq 6$ and with stellar masses $M_* \geq 10^9 M_\odot$ in the two observed bands (F200W and F356W) with a temperature parameter $h = 0.5$ (that controls the strength of penalties on hard negative pairs). We randomly split our dataset into a training and a test sample consisting of 1 100 and 138 galaxies, respectively. This translates into a training and a test dataset of 22 000 and 2 760 galaxy images, respectively. We train our algorithm with a batch size of $N = 550$ (i.e., half the number of galaxies in the training set) for 1 500 epochs in a GPU NVIDIA T4 Tensor Core with 16 GB of RAM. Random data augmentation is applied every 50 epochs to increase the variability during the training process.

3.3. *Visualization of the representation space*

We can hence analyze the properties of the representation space learned by the SimCLR framework presented in the previous section.

We start by visualizing how the TNG50 galaxy images, both with and without noise, are distributed in the representation space. Since the representation space for each galaxy image is encoded into 1 024 features, we perform a dimensionality reduction from the 1 024 features to a 2D space to facilitate the visualization and interpretation of this representation. For that purpose, we use the Uniform Manifold Approximation and Projection (UMAP; McInnes et al. 2018) method with standard initial parameters (*metric* = *euclidean*, *n_neighbors* = 15 and *min_dist* = 0.1). The UMAP algorithm seeks to learn the manifold structure of the input data and find a low-dimensional embedding that preserves the essential topological structure of that manifold. It is therefore a way to visualize in 2D the representations learned by the self-supervised network. Before applying the UMAP technique, we assume the same distance metric in the representation space as the one used to calculate the contrastive loss in the head projection space and, therefore, we normalize the representations with an L^2 norm such that the euclidean and cosine distances between representations are equivalent. The two coordinate axes in the UMAP representation do not have any precise physical meaning and are a combination of the 1 024 dimensions extracted by the contrastive learning setting.

It is important to emphasize that the contrastive approach is not intended for dimensionality reduction but for obtaining a robust representation of galaxy morphologies in a different space than images, which explains the high dimensionality of the representation space. Several works have shown indeed that the performance of contrastive learning increases with representations of higher dimension (e.g.,Chen et al. 2020a). The UMAP projection is used only for visualization purposes.

In Figure 4, we show random examples of galaxies in the F356W filter (both with and without noise) in the UMAP 2D space. Note that some stamps on the right-

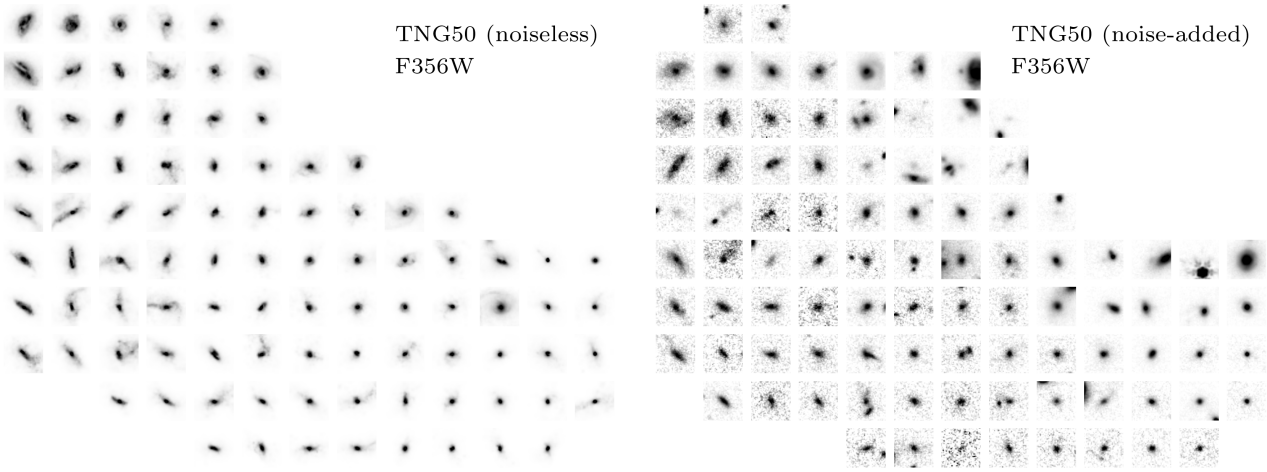

Figure 4. Randomly chosen images of the simulated TNG50 galaxies in the UMAP visualization. The UMAP space is binned and one galaxy image per bin is shown. The left panel shows the noiseless versions of the training galaxies, while the reduced versions (TNG50 + random CEERS patch) are shown in the right panel. Note that there is not a one-to-one correspondence between the galaxies shown in the two panels. Both panels correspond to galaxy images in the F356W filter.

hand panel (along with the addition of observed noise) show one (or more) foreground/background sources in the field of view. The figure clearly shows that galaxies are not randomly organized in the plane, indicating that the network has learned some morphological features. The distributions are also similar for galaxies with and without noise. Galaxies with extended light distributions and with clear signs of a disk component —or interactions— tend indeed to appear on the upper and upper-left parts of the UMAP space, while more compact galaxies with smoother and concentrated light distributions tend to be placed on the bottom-right section of the plane. We can also see that galaxies showing more elongated shapes are found on the left section of the representation space. Also interesting to notice is how several galaxies with bright companions (off-center sources) tend to be placed on the upper-right and bottom-right corners of the right-half panel (see subsubsection 3.4.1 below for a more detailed discussion on this point).

3.4. *A morphological description of galaxies robust to noise and background/foreground contaminants*

We now quantify in more detail the differences between the representations of noise-added and noiseless TNG50 galaxy images, and how the different augmentations of the same galaxy are represented by our contrastive model. As described in subsection 3.1, one of the reasons for using the SimCLR framework is to obtain a data-driven representation that is robust to noise and other observational effects such as foreground and background companions.

3.4.1. *Robustness to noise*

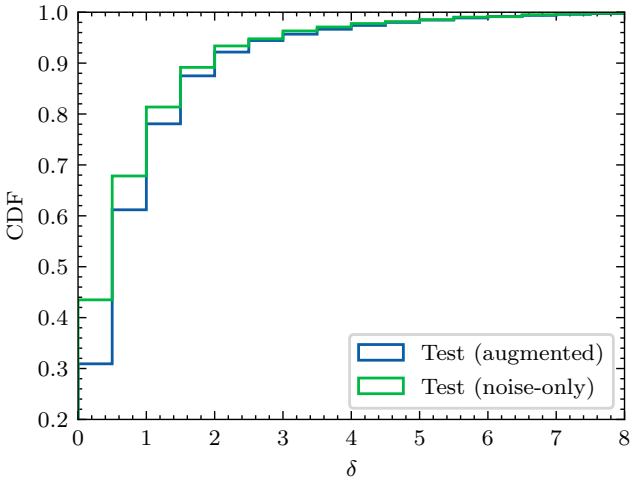

Figure 5. Cumulative distribution function of the distance in the UMAP plane between pairs of the same galaxy images, denoted as δ. Green histogram corresponds to the distribution of δ for the test dataset when only noise is added to the image, while the blue histogram shows the distribution of δ for the same dataset when all augmentations are applied to the noise-added image.

We first quantify the effect of noise by computing the distance in the UMAP representation between the noiseless and the noise-added images of each galaxy, denoted as δ. For a reference of the UMAP axis ranges, the horizontal axis (UMAP 1) spans within $(-1.9, 9.2)$ and the vertical axis (UMAP 2) spans within $(-0.9, 8.2)$. The total area covered by the data points in the UMAP plane is approximately 65 (in the arbitrary UMAP units, see Figure 6, for instance).

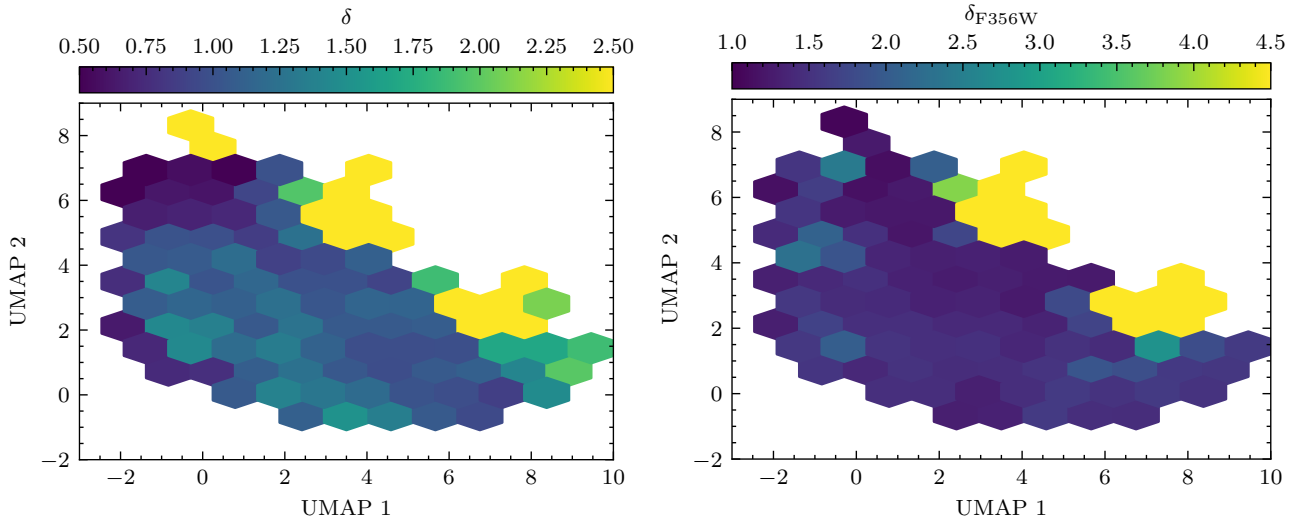

Figure 6. Left-hand panel: UMAP visualization for all the TNG50 galaxy images in our dataset color-coded by the mean value of the distance in the UMAP representation between the noiseless and the reduced images of each galaxy (denoted as δ). Right-hand panel: UMAP visualization for all the TNG50 galaxy images in our dataset color-coded by the mean value of the ratio of the flux measured in the TNG50 stamps and the flux derived from the noiseless TNG50 stamps for the F356W filter (denoted as $\delta_{\rm F356W}$). Large values of $\delta_{\rm F356W}$ indicate the presence of a secondary (or even more) source. The larger $\delta_{\rm F356W}$ is, the brighter the companions are with respect to the central galaxy.

In Figure 5, we show the cumulative distribution function of δ for the galaxy images in the test set for which only noise is added to the noiseless images. The rest of the augmentations (i.e., flip, rotation, and flux and color variation) are not applied. We find that 75% and 90% of the projections show values of $\delta \lesssim 1.3$ and $\delta \lesssim 2.1$ in the UMAP space, respectively. Also shown in the image is the cumulative distribution of δ for the same galaxy images with all the augmentations applied, as described in subsection 3.2. For this dataset, we find that 75% and 90% of the projections show values of $\delta \lesssim 1.4$ and $\delta \lesssim 2.2$ in the UMAP space, respectively. When augmentations are applied, the values of $\delta \lesssim 1$ are enlarged, but still remarkably well constrained within $\delta \lesssim 2$ for 90% of the dataset compared to the test set with no other augmentations applied besides noise. In other words, a value of $\delta \approx 2$ is analogous to saying that the noise and noiseless representations of the same galaxy pair are located within a circumference of radius $\delta/2 = 1$. Converted into an area, this means $\approx 5\%$ displacement in the UMAP plane for 90% of the galaxy images (i.e., $\delta \lesssim 2$) in the test set despite the level of noise, contamination and augmentations applied to the input galaxy images, as can be seen in Figure 3 and Figure 4.

We note also that the distribution of δ for the training and test samples are very similar. Therefore, we emphasize the model is not suffering from over-fitting, since none of the galaxy images in the test set has been shown previously to the network.

3.4.2. *Robustness to background/foreground contaminants*

We then check how our contrastive model behaves when more than one source (i.e., companions) is present in the stamp. To do so we measure the total flux in the noiseless TNG50 stamps (therefore, the intrinsic flux of the central galaxy) and in the TNG50 + random CEERS patch. The difference between both derived fluxes is a proxy for the presence (or not) of companions and, if present, how bright they are with respect to the central galaxy. In Figure 6, we show the UMAP plane for TNG50 galaxy images in our dataset color-coded by the distance in the UMAP δ and the ratio of the fluxes in the TNG50 stamps and the flux derived from the noiseless TNG50 stamps for the F356W filter (denoted as $\delta_{\rm F356W}$). It is interesting to note how the main two yellow clumps in the UMAP plane where the stamps with bright companions (more than three times the flux than the flux of the central galaxy, $\delta_{\rm F356W} \gtrsim 3$) tend to concentrate, and also their correlation with large values $\delta \gtrsim 2$. Some of these cases can be seen in the right-hand panel in Figure 4. For instance, there are several examples within these regions of $\delta_{\rm F356W} \gtrsim 3$ that correspond to TNG50 images for which the companion is so bright (such as a star) that the central galaxy cannot be even identified in the stamp. In these cases with bright companions around the central galaxy, the model tends to classify the brightest component (thus, the companion) instead of the central galaxy.

To further illustrate the effect of companions and noise on the representation space, we show some examples of

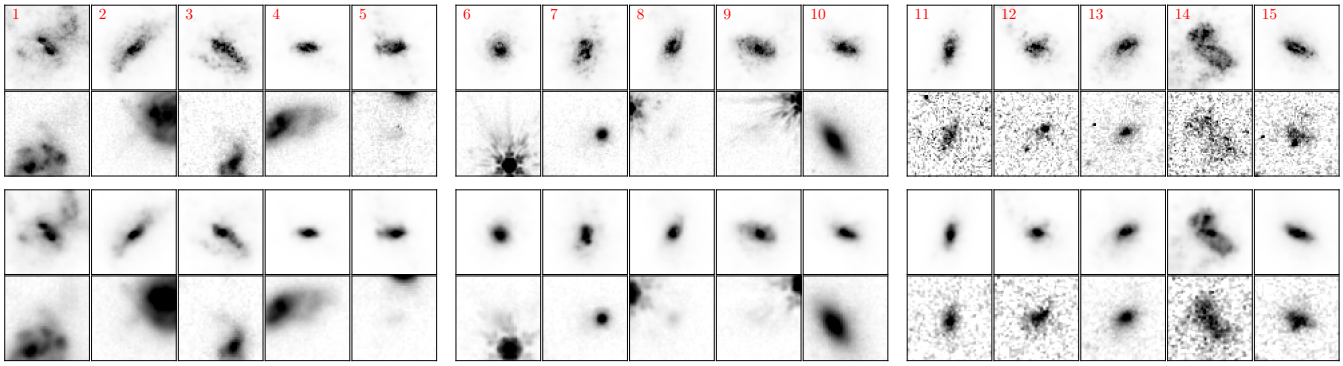

Figure 7. Randomly chosen examples of galaxy images with $\delta > 3$ (see Figure 6). For each of the examples, we show noiseless and noise-added images in the F200W (top rows) and F356W (bottom rows) filters. Cases 1 to 5 and cases 6 to 10 correspond to images with extended and compact bright companions, respectively. Cases 11 to 15 correspond to images with artifacts (such as cosmic rays). See subsubsection 3.4.2 for more details.

the most extreme cases ($\delta \gtrsim 3$) in Figure 7. The majority of images with the largest values of δ are mainly due to the presence of bright companions in the galaxy images (or artifacts). On one hand, in the noise-added images of cases 1 to 5 (belonging to the yellow clump with UMAP 2 $\gtrsim 4$ in the right-hand panel of Figure 6) it is possible to identify extended bright companions. On the other hand, in the noise-added images of cases 5 to 10 (belonging to the yellow clump with UMAP 1 $\gtrsim 6$ in the right-hand panel of Figure 6) it is possible to identify compact bright companions, mainly stars (cases 6,7 and 10). Finally, we also show 5 examples of galaxies (cases 10 to 15) belonging to the small yellow clump with UMAP 2 $\gtrsim 7$ in the left-hand panel of Figure 6. In the F200W noise-added images of these cases, it is possible to identify some burned pixels and saturated black patches that are due to cosmic rays or artifacts originated during the reduction process of the CEERS pointings used to add noise. Therefore, finding galaxy images that are located in that region of the UMAP is indicative of some quality image problems or contamination by other sources in the images.

Removing the cases with $\delta \gtrsim 3$ will even reduce the differences between the representations of the noise-only and the augmented datasets (Figure 5). Nevertheless, there is a significant fraction of TNG50 galaxy images with close companions that are still represented according to just the central galaxy in the stamps since they have values of $\delta \lesssim 2$.

Therefore, training our contrastive model with a combination of noiseless and noise-added TNG50 images leads to a robust representation of TNG50 images even in the case of the presence of companions in the image (at least, for those cases in which the companion is not extremely bright compared to the central galaxy). For the cases in which the companion is much brighter than the central galaxy, their locations in the UMAP may certainly help to find these cases in observed images and to treat them carefully in subsequent analysis.

3.5. *Dependence on physical and photometric parameters*

An advantage of calibrating the neural network model with simulations is that we have access to a large number of physical properties of the galaxies. An additional test for our classification scheme is, therefore, to examine how the representation space is correlated with physical quantities as well as with other (more standard) morphological measurements.

To increase the number of galaxies and given the considerations described in subsubsection 3.4.1, hereafter, we show representations and galaxy images for the whole dataset and remove those cases with $\delta > 3.0$ (corresponding to $\sim 10\%$ of the dataset). By doing so we retrieve a more reliable representation of the galaxies in our dataset without the impact of bright companions and artifacts.

3.5.1. *Correlation with physical properties*

In this section, we discuss how some physical properties extracted from the TNG50 simulation correlate with the representation in the UMAP plane. In Figure 8, we show the dependence in the UMAP plane with the total stellar mass ($M_*[\mathrm{M}_\odot]$), the specific angular momentum of stars ($j_*[\mathrm{kpc\ km\ s^{-1}}]$), the mass fraction in non-rotating stars (f_{nr}) and the flatness ($1 - f$) of the galaxy. The mass fraction in stars that have no net angular momentum around the z-axis is defined using the circularity parameter $\epsilon = J_z/J(E)$, as in Marinacci et al. (2014), for every star particle. It measures the maximum specific angular momentum possible at the specific binding energy E of the star. The mass fraction in non-rotating stars mass (denoted as f_{nr}) is then defined as the fractional mass of stars with $\epsilon < 0$ mul-

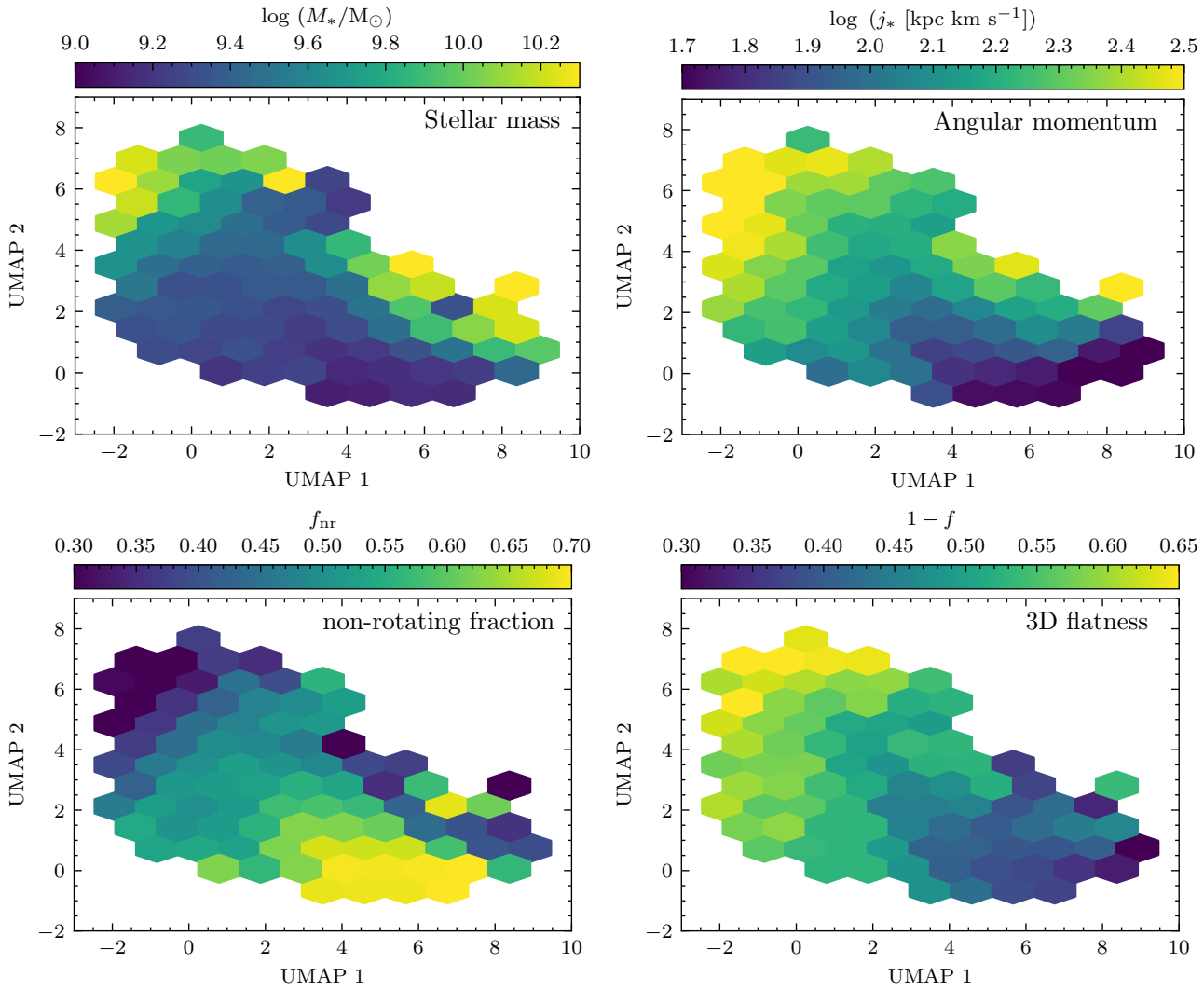

Figure 8. UMAP visualization for all the TNG50 galaxy images in our dataset color-coded by the distribution of several physical properties extracted from the TNG50 simulation. Color code corresponds to the median values in each hexagonal bin in the UMAP plane. From left to right and top to bottom, the different panels show: the logarithm of the total stellar mass ($\log M_*$[M$_\odot$]), the logarithm of the specific angular momentum of the stars ($\log j_*$[kpc km s^{-1}]), the mass fraction in non-rotating stars ($f_{\rm nr}$) and the galaxy flatness ($1-f$). The scatter maps of these parameters are presented in Appendix A.

tiplied by two. The flatness of the galaxy is computed as follows: $f = c/\sqrt{ba}$, where $c < b < a$ denote the principal axes obtained as the eigenvalues of the mass tensor of the stellar mass inside $2r_*$. The larger $1-f$ is, the flatter the system is in 3D. Here and throughout the paper we refer to the definitions and measurements of Pillepich et al. (2019). See also subsection 5.2 for a more detailed discussion of the 3D shapes of the TNG50 galaxies.

Figure 8 shows remarkable correlations between the position of galaxies in the UMAP and their average physical properties. Overall, galaxies with larger specific angular momentum and a flatter stellar distribution tend to populate the upper left region of the UMAP. These galaxies are also the most massive ones although the correlation is less clear. The bottom right corner is populated by rounder objects with lower specific angular momentum. It is also interesting to see that the transition between the variation of the physical properties is smooth, translating a continuum of galaxy morphology/structure.

Figure 8 only shows the median values of the physical properties in different regions of the UMAP. In order to quantify how constraining are these correlations, it is also important to measure the scatter of the different properties. This is shown in the appendix A (Figure A1). In most cases, the scatter represents less than $\sim 20\%$ of the dynamical range, indicating that the distributions are overall relatively narrow and, therefore, the correlations with physical properties are informative.

We conclude that the representation space for images —in addition to being robust to observational and instrumental effects— carries information about the kinematics and intrinsic shapes of galaxies.

3.5.2. *Connection to standard morphological measurements*

In Figure 9, we show the dependence on several photometric parameters estimated by Costantin et al. (2022b) in the F200W filter: the effective radius (r_e), the Sérsic index (n_e), the ellipticity from the Sèrsic fit ($1-b/a$), the concentration parameter (C), the asymmetry ($|A|$) and the smoothness (S). There is a remarkable correlation between the position in the UMAP plane and n_e, C, A and S. Gradually, n_e C and S grow from left to right in the UMAP space, while the A does it from right to left. Therefore, galaxy images with smoother, symmetric and concentrated light distributions are found towards the right section of the UMAP plane. Also important is the correlation with the ellipticity ($1-b/a$), with more elongated galaxies lying on the left (bottom-left) section of the UMAP plane. To illustrate again the spread of these representations, we show in the appendix A (Figure A2) the scatter of the parameters shown in Figure 9 as the standard deviation divided by the mean values in each hexagonal bin in the UMAP plane.

It is important to notice the existing correlation with the physical effective radius, r_e, with the largest galaxies populating the top-left section of the UMAP plane. This correlation with the physical size is not in contradiction with the representations being independent of apparent galaxy sizes but reflects instead the known correlation between morphological appearance and physical size (e.g.,van der Wel et al. 2014a).

Based on the previous maps calibrated with the TNG50 simulation, asymmetric, more extended, flatter and rotationally-supported galaxies tend to populate the upper-left section of the UMAP representation. In more detail, the more to the left in the UMAP plane a galaxy is, the more elongated it appears. Smoother, more compact, rounder and non-rotating galaxies are located toward the bottom-right corner of the UMAP representation. Although not shown, the results presented here are consistent (despite small variations) for the same morphological parameters measured in the F365W filter.

4. SELF-SUPERVISED LEARNING REPRESENTATION OF JWST GALAXY IMAGES

In this section, we apply the methodology described before to the two datasets of observed galaxies with JWST described in subsection 2.1.

4.1. *Representations of CEERS galaxy images*

We feed the 891 observed CEERS galaxies to our contrastive model to retrieve their corresponding representations in the 1 024 dimensions space. Then, we normalize the derived features and transform them into a 2D vector using the same UMAP embedding obtained for the features of the TNG50 galaxy images.

In the top row of Figure 10, we show a comparison of the UMAP representation space for our initial (TNG50 + random CEERS patch) and the observed CEERS datasets. Interestingly, observed galaxies tend to populate the complete UMAP plane which indicates that both samples share similar morphological diversity. The UMAP visualization is however a projection of a higher dimension space, which is not appropriate for outlier detection. Even if observed galaxies would not reside in the same manifold as simulated objects, the UMAP representation would tend to show them towards the edges of the plane but not outside. This is the behavior seen for observed CEERS galaxies which tend to be concentrated in the edges of the UMAP cloud —towards the bottom and bottom-right sections— independently of the source redshift. Given that the mass and flux distributions of both datasets are consistent —even though we have not performed a careful one-to-one match between simulations and observations— the differences in the distributions of points of both datasets are likely to originate in intrinsic differences in the morphological properties. Combining the distributions of points in Figure 10 with the information provided by Figure 8 and Figure 9, we conclude that observed CEERS galaxies occupy more frequently than the simulated TNG50 galaxies the regions in the representation space where galaxies are more compact and with less specific angular momentum. We investigate these differences in more detail in section 5.

4.2. *Representations of VISUAL galaxy images*

We also present a comparison of the representation obtained after applying our contrastive model to the VISUAL dataset for which visual morphological classifications are provided (Kartaltepe et al. 2022). After selecting those galaxies with $M_* \geq 10^9 M_\odot$, $3 < z < 6$ and reliable visual classifications we end up with a dataset of 523 galaxies. From this dataset, there are 307 candidates ($\sim 59\%$) classified as disks in four categories: 117 *Disk*, 103 *Disk+Irr*, 27 *Disk+Sph+Irr* and 60 *Disk+Sph*. On the other hand, 202 galaxies ($\sim 38\%$) are classified as spheroids in three categories: 92 *Sph*, 23 *Sph+Irr* and , 27 *Disk+Sph+Irr* and 60 *Disk+Sph*. Finally, there are 82 galaxies classified as *Irr* ($\sim 16\%$), 4 as *point sources* and 2 as *unclassifiable*.

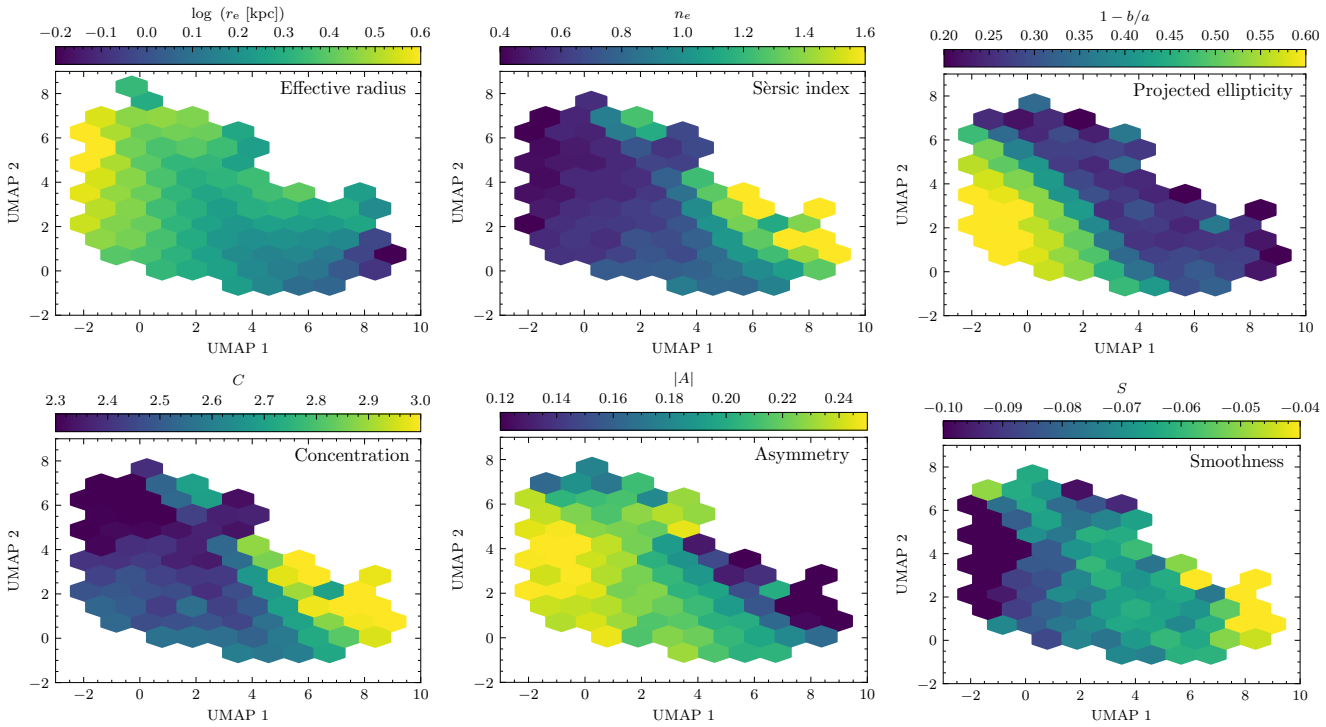

Figure 9. UMAP visualization for all the TNG50 galaxy images in our dataset color-coded by the distribution of several morphological and photometric parameters. Color code corresponds to the median values in each hexagonal bin in the UMAP plane. From left to right and top to bottom, the different panels show: the logarithm of the effective radius (r_e[kpc]), in kpc), the Sèrsic index (n_e), the ellipticity based on Sèrsic fit ($1 - b/a$), the concentration (C), the asymmetry (A) and the smoothness (C). The scatter maps of these parameters are presented in Appendix A.

In the bottom row of Figure 10, we show the representation of the galaxy in the VISUAL dataset in the UMAP plane for the various morphological groups based on the provided visual classifications. Although the mass and redshift selection of the galaxies is based on different estimators (JWST photometry for the CEERS dataset and CANDELS photometry for the VISUAL dataset), the distribution of the representations in the UMAP plane for the VISUAL galaxies is similar to the CEERS representations (i.e., a significant fraction of galaxies occupy the bottom section of the UMAP plane).

The figure reveals some expected correlations with the traditional visual morphology. It is reassuring that Disk+Spheroid and Spheroid groups from the VISUAL catalog populate the bottom-right section of the UMAP plane, where compact, non-rotating galaxies with low angular momentum (according to TNG50 properties) are expected to be. However, we notice that galaxies classified as Disk, Irregular and/or Disk+Irregular in VISUAL are distributed throughout the plane even towards the bottom-right section of the UMAP very close to where spheroids lie. As shown in Figure 8 and Figure 9, this lower right region of the UMAP where, according to VISUAL, disk-like morphologies are located corresponds to galaxies in TNG50 with physical and photometric properties typically shared by spheroidal systems, such as low specific angular momentum, large mass fractions in a non-rotating component, low flatness and large Sérsic indexes. This rises interesting questions about the true nature of these disks that we discuss in section 5.

5. IMPLICATIONS AND DISCUSSION: THE OPTICAL REST-FRAME MORPHOLOGIES OF HIGH REDSHIFT GALAXIES

In this section, we examine in more detail the differences found in previous sections between the simulated TNG50 and the observed JWST galaxy images. We also discuss the implications of our results for the nature of visually identified disks at $z > 3$ in the real Universe.

5.1. *Does the TNG50 model reproduce the morphologies of observed galaxies?*

5.1.1. *Distribution of representations*

The representations of the simulated TNG50 and the observed CEERS galaxy images inferred by our contrastive model seem to be distributed differently (subsection 4.1 and subsection 4.2). Observed CEERS galaxies tend to concentrate in the bottom section of the UMAP plane, while simulated TNG50 galaxies expand

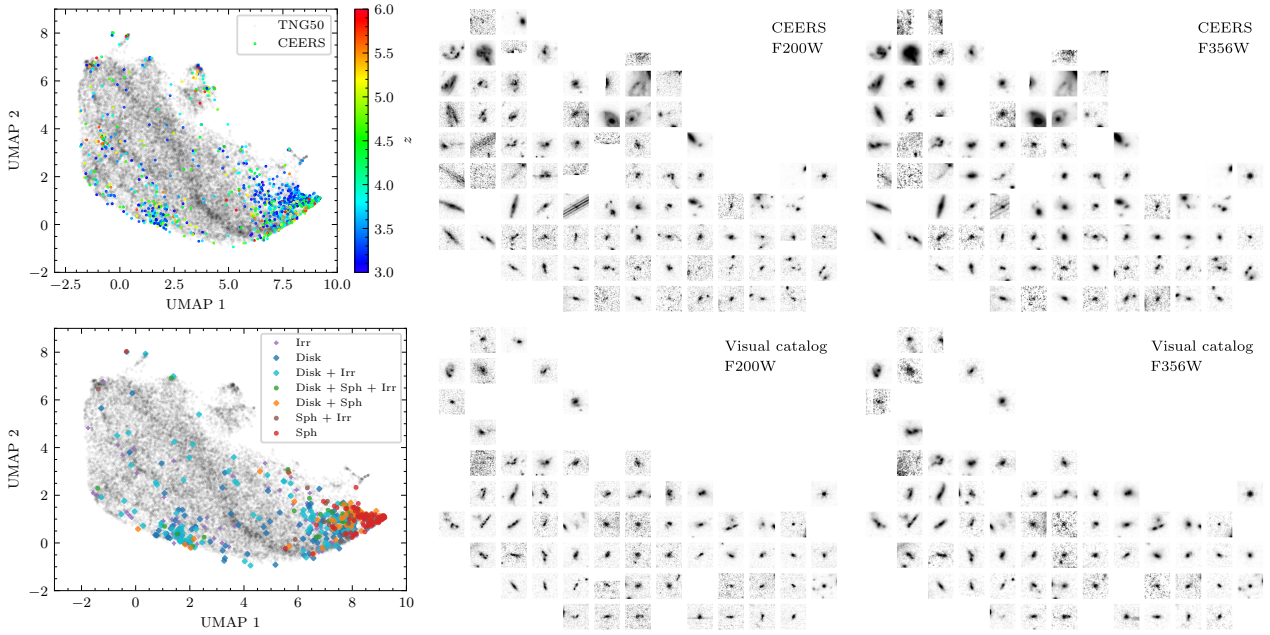

Figure 10. Comparison of distributions of observed and simulated galaxies in the representation space. The top row shows the CEERS mass-complete sample and the bottom row the VISUAL sample with available visual morphologies. Top left-hand panel: UMAP visualization for the observed CEERS galaxy images selected in mass and redshift (color-coded by the source photo-z) overlapped with the representation of noise-added TNG50 galaxy images. Top middle panel: randomly chosen observed CEERS galaxy images in the UMAP visualization in the F200W filter. Top right-hand panel: randomly chosen observed CEERS galaxy images in the UMAP visualization in the F356W filter. Bottom left-hand panel: UMAP visualization for the observed VISUAL galaxy images selected in mass and redshift. Points are colored according to the visual classifications into several non-exclusive classes. Bottom middle panel: randomly chosen observed VISUAL galaxy images in the UMAP visualization in the F200W filter. Bottom right-hand panel: randomly chosen observed VISUAL galaxy images in the UMAP visualization in the F356W filter.

over the whole UMAP range with similar number densities. As previously mentioned, the UMAP representation is not well suited for the detection of outliers. Therefore, even if observed galaxies seem to overall lie in the same region as simulated ones, they can still live in different manifolds in the higher dimensionality representation space.

To further quantify this distribution shift, we first derive the distance - in the 1 024 dimensionality space - to the 10th closest TNG50 neighbor for each galaxy in the VISUAL and CEERS datasets (δ_{10}). In order to have a fair reference distribution, we select, for each observed galaxy, the closest simulated one in the representation space and compute also δ_{10}. If both datasets - observed and simulated - live in the same manifold the distribution of distances should be similar. If on the contrary, observed galaxies populate different regions of the parameter space, their representations should be more isolated and therefore we should measure larger values of δ_{10}. The distribution of distances δ_{10} is shown in Figure 11. We clearly see that the distributions for observed galaxies are shifted towards larger values compared to the reference distribution. This indicates that observed galaxies are on average more isolated than simulated ones in the representation space. This separation could be interpreted as an additional indication that the representations obtained for the TNG50 and the JWST observations do not exactly live in the same manifold.

To better understand these measured discrepancies, we quantify the differences between observed and simulated galaxies in terms of more standard morphological properties in Figure 12. We show the distributions of observed and simulated galaxies in the $\log M_* - \log r_e$, $\log M_* - \log n_e$ and $\log M_* - b/a$ planes in four redshift bins. To divide in redshift, we take all galaxies in a given snapshot for the simulation and observations are associated with the closest snapshot based on the photometric redshift. It should be kept in mind that this figure (as the previous ones) do not include all TNG50 galaxies, but those for which the JWST mocks are available for a field-of-view larger than 64×64 pixels (see subsection 2.2 and Figure 2 for more details). Given the small amount of galaxies removed, we do not expect the distributions to change significantly though.

It is manifest, firstly, that the TNG50 simulated galaxies overlap with CEERS observed ones in the parame-

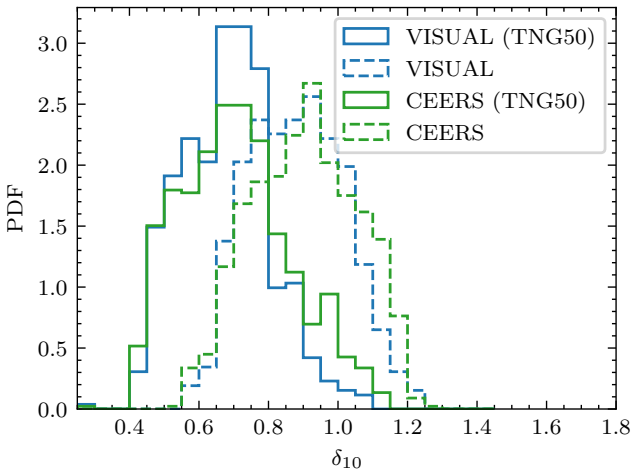

Figure 11. Probability density functions of the distances to the 10th closest neighbor in the 1 024 dimensions of the representation space, denoted as δ_{10}. Solid histograms correspond to the distance to the 10th closest neighbor in TNG50 of the closest TNG50 neighbor of each galaxy in the VISUAL (in blue) and the CEERS (in green) datasets. Dashed histograms correspond to the distance to the 10th closest neighbor in TNG50 of each galaxy in the VISUAL (in blue) and the CEERS (in green) datasets.

ter spaces of Figure 12. This is per se, again, a non-negligible confirmation of the zeroth-order good functioning of the underlying model. However, it is also apparent, differently than what could be deduced from the representation space distributions, that the TNG50 galaxies studied here actually exhibit less galaxy-to-galaxy variation in sizes, Sersic indices and shapes than CEERS observed galaxies, at fixed stellar mass and redshift. Furthermore, the TNG50 simulation predicts galaxies with larger sizes (at $z = 3{-}4$, but not $z = 5{-}6$), with smaller values of n_e (at all $z = 3 - 6$) and that are rounder in projection (more so the higher the redshift) than what is measured in CEERS. These differences at least partly explain the different distributions in the representation space of contrastive learning and also go in the expected direction of observed galaxies mainly populating the bottom right corner of the UMAP.

These reported differences could originate from a resolution-induced effect (see e.g., Zanisi et al. 2021) or could be an indication of more fundamental physical differences. We note as well that the stellar masses reported for the TNG50 simulation correspond to the 3D stellar mass, while those obtained for the CEERS dataset are based on the SED fitting to the JWST photometry. Also, the Sersic parameters for the TNG50 and the CEERS galaxies are derived using different methodologies: for the TNG50, the morphological parameters are obtained with `statmorph` (as described in Costantin et al. 2022b); while for the CEERS dataset, they are derived with `galfit`. More in-depth comparisons of simulated and observed data—likely beyond images—are required to reach a final conclusion.

5.1.2. *Morphological classes*

As an additional way to quantify the differences between TNG50 galaxies and observed CEERS galaxies, we compare the abundances of TNG50 and CEERS galaxies retrieved from the separation into two main classes (for simplicity) using a clustering technique. In particular, we apply the $k-$means algorithm to cluster data in the representation space by trying to separate samples in k groups of equal variance, minimizing a criterion known as the inertia or within-cluster sum-of-squares (WCSS).

In Figure 13, we show the UMAP visualization color-coded by the two classes for the simulated TNG50 and the observed CEERS datasets. Note that galaxies with artifacts or bright companions are not included in the derivation of the different class fractions hereafter, although they are shown in the UMAP visualization panel. The clustering algorithm naturally separates the upper and lower regions of the UMAP. Given the division, we denote those galaxies located in the upper-left section of the UMAP as *extended* (E) galaxies, while those located in the bottom-right section as *compact* (C) galaxies. We find that $\sim 52\%$ and $\sim 48\%$ of the galaxies belong to the E and C class, respectively. To reinforce the conclusion that the model is robust to noise, we find a 92% agreement between the two classes for the noiseless and the noise-added TNG50 dataset. This can be interpreted as classes being consistent independently of the galaxy image shown to the contrastive model, i.e., the noiseless or the noise-added version.

The fraction of galaxies belonging to the *compact* (C) class as a function of the stellar mass for both observed and simulated galaxies confirms that the TNG50 model systematically under-predicts the abundance of compact galaxies.

5.2. *Are visually classified disks really disky?*

We now combine the visual classifications with the positions of galaxies in the representation space of contrastive learning—which have been shown to correlate with physical properties—to try to establish new constraints on the abundance of disks at $z > 3$. As shown in section 4.1, visually classified disks are spread all over the representation space which suggests that disks represent a heterogeneous group of galaxies with different physical properties.

5.2.1. *3D shapes from the representation space*

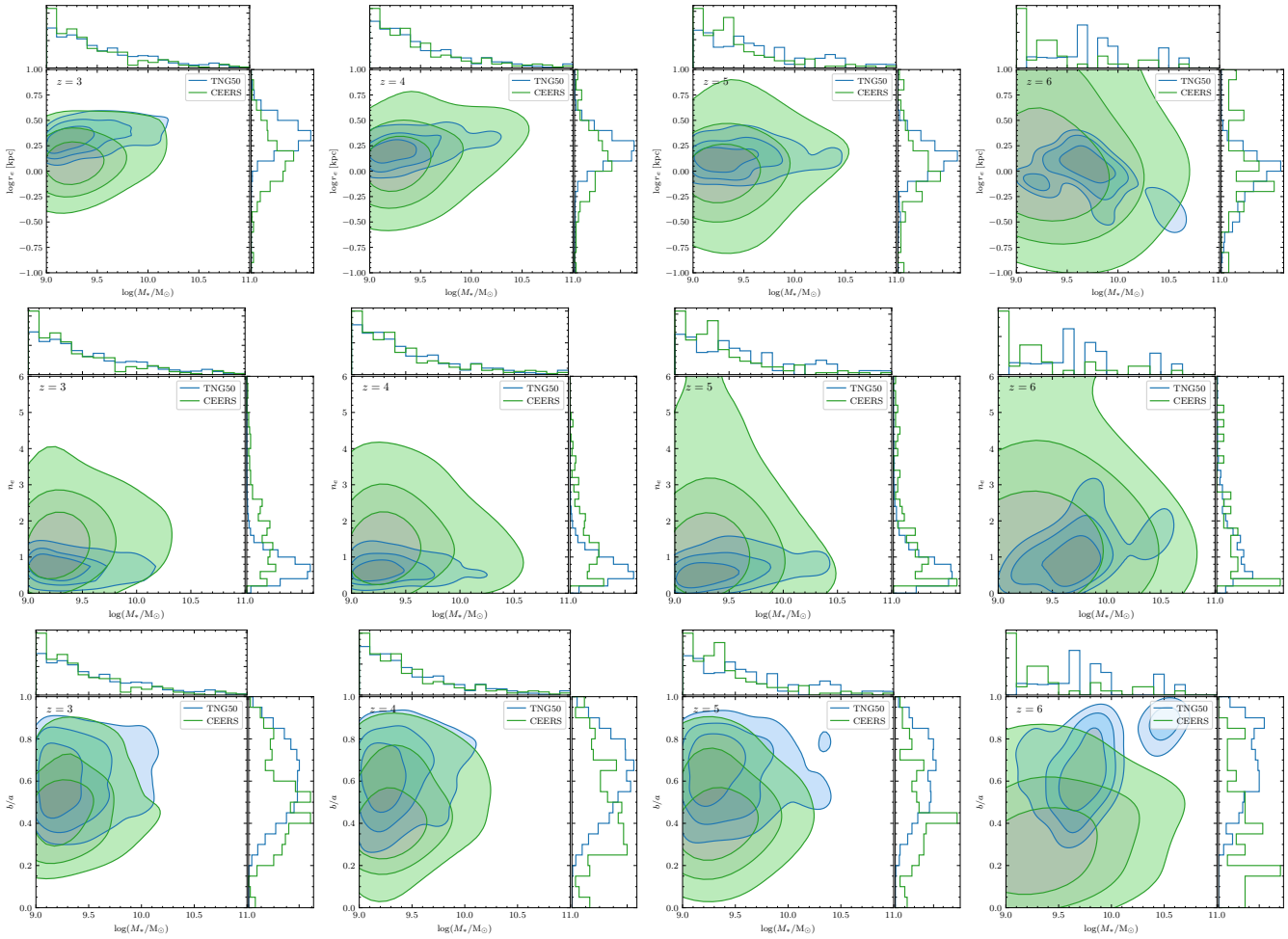

Figure 12. Distribution of the logarithm of the effective radius ($\log r_e$ in kpc, top row), the Sèrsic index (n_e, middle row), and the axis ratio (b/a, bottom row) as a function of the logarithm of the stellar mass ($\log M_*$) for the TNG50 (in blue) and the CEERS (in green) datasets. From left to right, the panels show the distributions at $z = 3, 4, 5, 6$ for the TNG50 dataset. For the CEERS dataset, galaxies are included in the closest redshift value. Contour levels enclose 25%, 50% and 75% of the data. The photometric parameters shown are measured in the F200W filter.

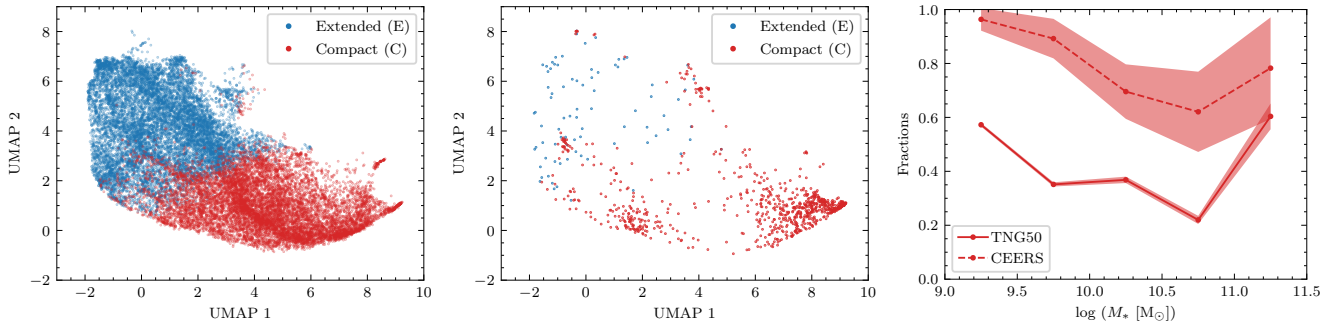

Figure 13. Left-hand panel: UMAP visualization of noise-added TNG50 galaxy images color-coded by classes according to the k-means method for two clusters: *Extended* (E) class (in blue) and *Compact* (C) class (in red). Middle panel: same as left-hand panel but for the CEERS galaxy images. Right-hand panel: fractions *Compact* (C) galaxies in TNG50 (solid lines) and CEERS (dashed lines) in 5 logarithmic mass bins of width 0.5 dex in the range $9 <= \log(M/\mathrm{M}_\odot) <= 11.5$. The shaded regions correspond to the fraction errors considering Poisson errors in the number of selected galaxies and the total number of galaxies in each mass bin.

We start by exploring in more detail how the contrastive learning representation space distributes galaxies with different 3D structures of the stellar mass dis-

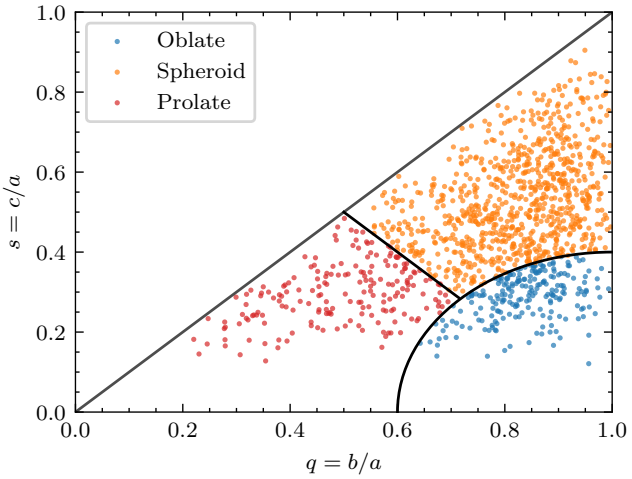

Figure 14. Minor-to-major (s) versus intermediate-to-major (q) axis ratios of the stellar mass distribution for the TNG50 galaxies analyzed in this work. Galaxy shapes are split into *Oblate* (in blue), *Spheroid* (in orange) and *Prolate* (in red) according to the definition of van der Wel et al. (2014b) and the color code used in Pillepich et al. (2019).

tribution. We characterize the shape of galaxies in the TNG50 sample as done and studied in Pillepich et al. (2019), i.e., with an ellipsoid with three semi-axes $c < b < a$ and use the axial ratios $q = b/a$ (intermediate-to-major) and $s = c/a$ (minor-to major) to define three main 3D shape classes: oblate, spheroid and prolate following the definitions of van der Wel et al. (2014b) and Zhang et al. (2019). The axial ratios are derived after diagonalizing the stellar mass tensor in an iterative way while keeping the major axis length fixed to $2r_*$. We consider oblate or disky galaxies those with $a \sim b > c$, elongated or prolate objects those with $a > b \sim c$ and, finally, spheroidal systems those with similar values for the three semi-axes. Figure 14 shows the distribution of the TNG50 sample used in this work ($z > 3$ and $\log M_*/M_\odot > 9$) in the $s - q$ plane used by van der Wel et al. (2014b) to define the three main 3D shapes. Note that, by definition, the 20 projections of the TNG50 galaxies have the same 3D shape. At these redshifts, we find that $\sim 67\%$ of the galaxies in the simulation have a spheroidal shape (828 out of 1238) according to this definition. Only 217 (17%) and 193 (16%) have oblate and prolate shapes, respectively. The fact that at high redshift and low stellar masses, galaxies tend to present a more prolate structure has been found both in observations (van der Wel et al. 2014b; Zhang et al. 2019) and simulations (Tomassetti et al. 2016; Pillepich et al. 2019).

We explore in Figure 15 the distribution of the mock images of prolate, oblate and spheroid galaxies in the UMAP projection of the representation space obtained with contrastive learning. We consider for this exercise the 20 projections of the same galaxy as independent objects which explains the larger number of objects compared to Figure 14. Interestingly, disky galaxies tend to be located roughly in the upper-left half section of the UMAP manifold, while elongated systems are distributed all over the UMAP plane. Although spheroidal galaxies populate all the plane because they vastly dominate in numbers, the results of Figure 15 suggest that galaxies located in the bottom-right corner of the UMAP (i.e., outside the 95% contour level) are very unlikely to be disky even if they appear as elongated in the images. However, Figure 10 shows that visually classified disks are distributed all over the place and that only a small fraction is located in the region where oblate systems are expected to be. It suggests again that not all the visually classified disks have the same properties.

5.2.2. *Oblate systems*

Based on the previous result, we select galaxies with a disk visual morphology and a high probability of being oblate in shape to further study their properties. These systems should be considered as *true* flat disk candidates. By disk visual morphology we denote all galaxies belonging to one of the following classes: *Disk*, *Disk + Irr*, *Disk + Sph + Irr* and *Disk + Sph*. We then define as *oblate disk* candidates those galaxies with a disk visual classification which are located within the 95% contour for the oblate systems in the TNG50 dataset (see Figure 15). We note that only one galaxy classified as *Sph* and two classified as *Sph + Irr* fall within the oblate contour defined that way. We find 128 *oblate disk* candidates out of 307 galaxies visually classified as disks, which corresponds to $\sim 42\%$ of the galaxies with a disk visual morphology. In detail, from the 128 candidates for *oblate disk*: 50 galaxies are visually classified as *Disk*, 8 galaxies are classified as *Disk + Sph*, 61 galaxies are classified as *Disk + Irr* and 9 galaxies as *Disk + Sph + Irr*. In Figure B1, we show the galaxy images of the *oblate disk* candidates. They all present extended light distributions with, in many cases, signs of internal structure or irregularities.

5.2.3. *Non-oblate systems*

Following a similar approach, we now turn our attention towards the objects that, according to our representation, are very unlikely to be *true* oblate disks, but are visually classified as such. This should allow a better understanding of the morphological properties that drive the visual classification into disks. We focus on those galaxies visually classified as disks (including all the different disk categories) that are located outside the

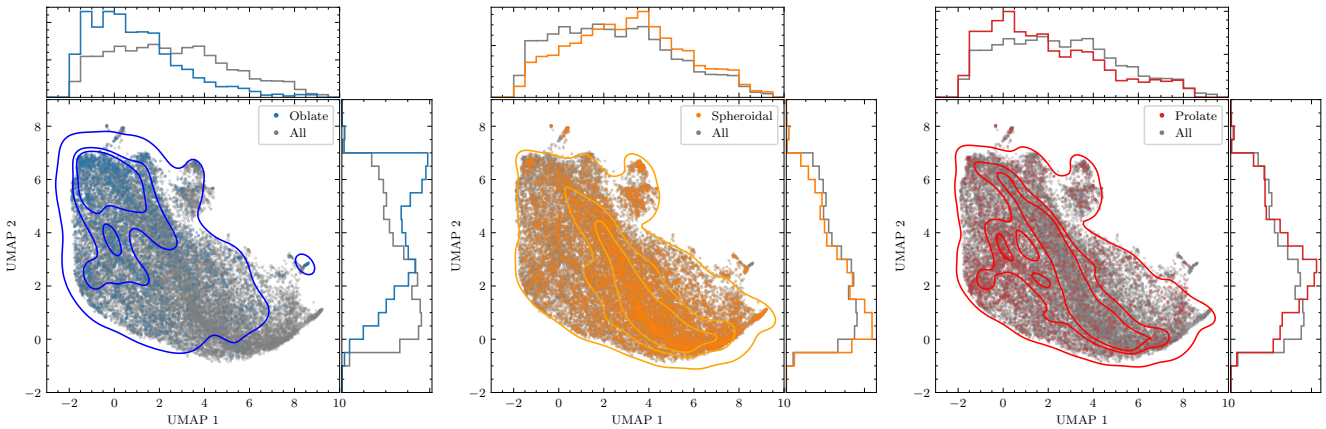

Figure 15. Location in the UMAP plane of TNG50 oblate (blue points in the left-hand panel), spheroid (orange points in the middle panel) and prolate (red points in the right-hand panel) galaxies according to the 3D shape inferred from the stellar particles. Morphologically-disky (i.e., oblate) galaxies tend to populate the upper regions of the UMAP plane, while prolate and spheroid systems cover the whole UMAP space. The contour levels indicate the 25%, 50% and 95% probabilities

95% probability contour for oblate systems. Hereafter, we denote these galaxies as *non-oblate disk* galaxies. We find that $\sim 58\%$ (179 out of the 307) of the galaxies with disk visual morphologies fulfill this selection criterion. In particular, out of these 179 *non-oblate disk* candidates, we find that 67 are classified as *Disk*, 52 are classified as *Disk + Sph*, 42 are classified as *Disk + Irr* and 18 as *Disk + Sph + Irr*. On the other hand, among the *non-oblate* systems, 91 galaxies are classified as *Sph*, 21 are classified as *Sph + Irr* and 30 are classified as *Irr*.

In Figure B2, we show some examples of *non-oblate* galaxies that are visually classified as *Disk*, *Disk+Irr* and *Irr*, along with some examples of *Sph* for comparison. Most of the disk galaxies present a more elongated shape in the 2D image compared to the spheroid galaxies which can be interpreted as a disk signature. This elongation is most likely driving the visual classification toward disk-like morphologies. In fact, the example images of galaxies lying in a similar region of the UMAP but visually classified as *Sph* are, on average, rounder than the ones visually classified as *Disk* or *Disk+Irr*.

5.2.4. *Two populations of visually classified disks*

The results from the previous subsections suggest that there are two different populations of visually classified disks which we call *oblate disk* and *non-oblate disk* candidates. This result relies somehow on the calibration of the representation space with the TNG50 simulation which, as shown in subsection 5.1, might introduce some unknown biases because the morphologies of observed and simulated galaxies do not perfectly match. Nevertheless, it shows that prolate or spheroidal systems might appear elongated in noise-added 2D images and, therefore, be misidentified as disks. This effect should be independent of the degree of agreement between simulated and observed galaxies.

We now examine the properties of the two populations using standard projected shape measurements in a similar way as done in Zhang et al. (2019). In Figure 16, we show the distribution in the axis-ratio/semi-major axis plane ($b/a - \log a$) of *oblate disk* and *non-oblate disk* candidates (along with *non-oblate spheroids* galaxies for comparison) according to the contrastive learning approach. We denote by *non-oblate spheroid* galaxies those objects visually classified as *Sph* and located outside the 95% probability contour for oblate systems. The b/a and a values in Figure 16 are derived from Sersic fits to the F356W light profiles (see section 2 for more details).

To make this comparison, we remove from the oblate class some objects which cluster in the UMAP plane at $0.5 \lesssim$ UMAP $1 \lesssim 2$ and $-0.5 \lesssim$ UMAP $2 \lesssim 1$ in Figure 10. Indeed, after a visual inspection, many of them show hints of a double-nucleus, interacting galaxies or galaxies with a close companion comparable in brightness to the central galaxy, which are therefore very unlikely to be disks. It is interesting though that these systems are located in a very specific region of the representation space. In Figure B3 and Figure B4 in Appendix B, we show 20 randomly chosen examples of these galaxies. They also look elongated in projection.

The Figure 10 shows that *oblate disk* candidates present a relatively flat distribution of b/a, as one would expect from a pure projection effect. They also have larger semi-major axes than the average population (peaking at $\log a \sim 0.2 - 0.3$), as also expected for disk galaxies. However, the *non-oblate disk* candidates are more concentrated towards smaller values of $b/a \lesssim 0.6$ and smaller semi-major axis (peaking at $\log a \sim 0.0$).

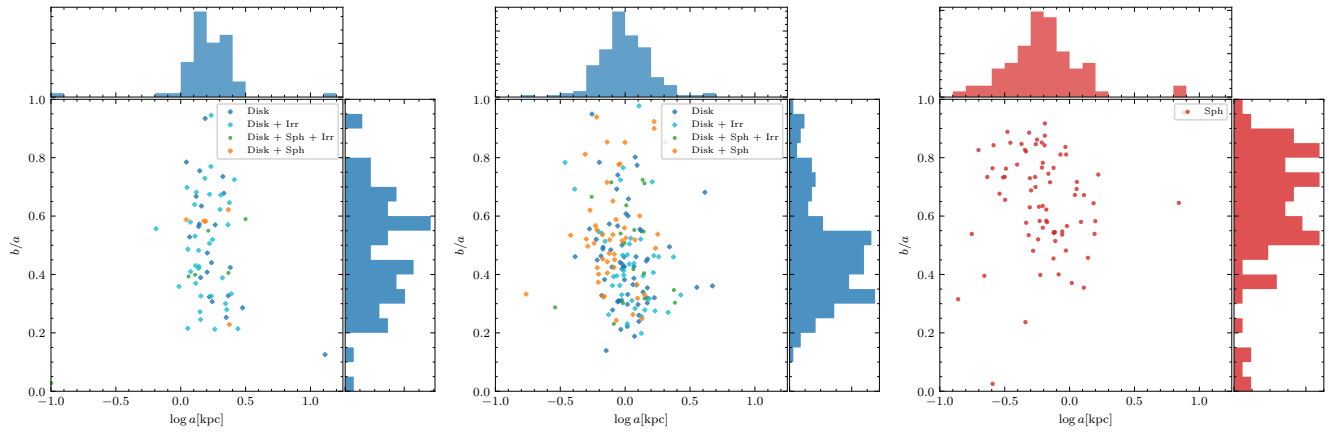

Figure 16. The projected $b/a-\log a$ distribution of CEERS galaxies observed with JWST and with available visual classification for: *oblate disk* candidates (left-hand panel), *non-oblate disk* candidates (middle panel) and *non-oblate spheroid* candidates (right-hand panel). All quantities are measured in the F356W filter. Note that cases falling in the region where double-nucleus and/or interacting galaxy are not included in the *oblate disk* candidates.

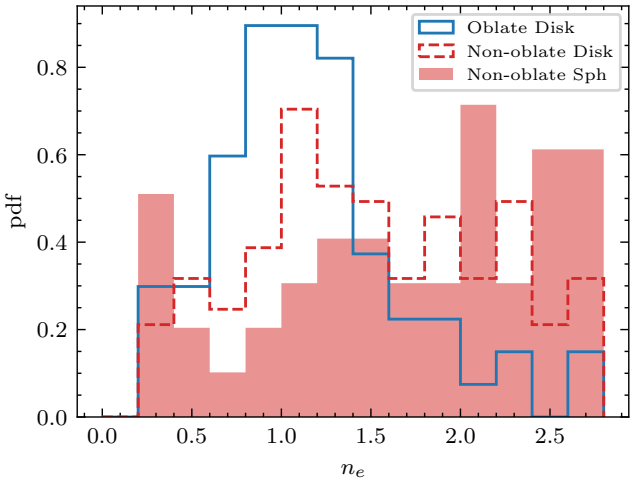

Figure 17. Probability density functions of Sèrsic index measured in the F356W filter for *oblate disk* (solid blue histogram), *non-oblate disk* (dashed red histogram) and *non-oblate spheroid* candidates (shaded red histogram) in the VISUAL dataset of JWST observed galaxies.

The lack of round systems in this population is also an indication that their elongated projected shapes are not only a projection effect but it is more a consequence of their intrinsic shape. This confirms the impression that these galaxies are visually classified as *Disk* and *Irr* mostly due to their elongated projected shapes, even if their intrinsic shapes are more likely to be prolate or spheroidal. For comparison, spheroids do clearly populate the region of high b/a values,

In Figure 17, we finally show the distribution of n_e for *oblate disk*, *non-oblate disk* and *non-oblate spheroid* candidates. It is clear how the distribution of n_e for the *oblate disk* candidates peaks at $n_e \lesssim 1$, characteristic of an exponential profile. Contrarily, the n_e distributions for *non-oblate disk* and *non-oblate spheroid* candidates are more skewed towards larger values of the Sèrsic index. This result is again indicative of the different properties of galaxies *Disk* visual morphologies.

Our results are also in qualitative agreement with the predictions of zoom-in cosmological simulations with fewer systems but somewhat better resolution and different underlying galaxy formation models (e.g. Tomassetti et al. 2016) which have measured a mass-dependent transition between prolate and oblate stellar shapes at high redshift. A more in-depth comparison of the physical properties of these populations and the predictions from different galaxy formation models is out of the scope of this paper, but should definitely provide additional clues.

6. SUMMARY AND CONCLUSIONS

This work presents a novel data-driven method based on contrastive learning to infer the morphological properties of galaxies observed with JWST. The method is calibrated on mock JWST galaxy images extracted from the TNG50 cosmological simulation that, thanks to its large number of qualitatively-realistic galaxies, allows us to produce a morphological description —without any assumption on galaxy classes— robust to noise, galaxy size, color and orientation. In addition to the robustness to noise, we show that the obtained representations of galaxies based on their images correlate well with some other physical properties inferred from the simulation (such as the specific angular momentum of stars, j_*, and the intrinsic 3D shape) along with some measured photometric and structural properties (such as Sersic index and ellipticity).

We have applied the method to JWST images from the CEERS survey in the F200W and F356W bands of:

1) a mass-complete sample ($M_* \geq 10^9 M_\odot$) of galaxies at $3 < z < 6$ in the CEERS survey; and 2) a mass- and a redshift-selected sample of CEERS galaxies at $3 < z < 6$ with $M_* \geq 10^9 M_\odot$ for which visual morphological classifications are available.

Our main results are:

- Simulated galaxies from the TNG50 cosmological simulation seem to cover well the observed morphological diversity at $z > 3$. However, the morphological distributions of CEERS and simulated galaxies are measured to be different. When compared at the pixel level, simulated and observed galaxies seem to populate in different proportions the different regions of the TNG50-trained manifolds. We show that these differences can be at least partly explained because observed galaxies can be more compact and more elongated than simulated ones. In fact, the galaxy-to-galaxy variation in sizes, Sersic indices and shapes at fixed stellar mass and redshift is larger in the observed CEERS population than in TNG50 simulated ones.

- Our morphological description also suggests that visually classified disks comprise two different populations: one made of *true* flat disks and another more compatible with having a prolate or spheroidal shape. A significant fraction of galaxies ($\sim 58\%$) that are visually classified as disks are indeed located in the representation space very close to compact spheroids and therefore are more consistent with having a prolate or spheroidal stellar structure. Although some of these conclusions are affected by the calibration with the TNG50 model, our study robustly confirms that some objects with a prolate or spheroidal intrinsic shape are elongated in the images and can be misclassified as disks. The coexistence of prolate and oblate systems at high redshift is in qualitative agreement with the predictions of other models (e.g. zoom-in simulations) which also found that low-mass galaxies at high-z tend to present a prolate shape (Ceverino et al. 2015; Tomassetti et al. 2016). More in-depth follow-up of these two populations of galaxies, possibly with spectroscopy, is required to further their true nature.

ACKNOWLEDGMENTS

MHC thanks Shy Genel, David Koo and Sandy Faber for insightful discussions. JVF, MHC, RS and JHK acknowledge financial support from the State Research Agency (AEIMCINN) of the Spanish Ministry of Science and Innovation under the grant "Galaxy Evolution with Artificial Intelligence" with reference PGC2018-100852-A-I00 and under the grant "The structure and evolution of galaxies and their central regions" with reference PID2019-105602GB-I00/10.13039/501100011033, from the ACIISI, Consejería de Economía, Conocimiento y Empleo del Gobierno de Canarias and the European Regional Development Fund (ERDF) under grants with reference PROID2020010057 and PROID2021010044, and from IAC projects P/300724 and P/301802, financed by the Ministry of Science and Innovation, through the State Budget and by the Canary Islands Department of Economy, Knowledge and Employment, through the Regional Budget of the Autonomous Community. JVF and FB also acknowledge support from the grant "Galactic Edges and Euclid in the Low Surface Brightness Era (GEELSBE)" with reference PID2020-116188GA-I00 financed by the Spanish Ministry of Science and Innovation. LC acknowledges financial support from Comunidad de Madrid under Atracción de Talento grant 2018-T2/TIC-11612 and Spanish Ministerio de Ciencia e Innovación MCIN/AEI/10.13039/501100011033 through grant PGC2018-093499-B-I00.

REFERENCES

Abraham, R. G., van den Bergh, S., Glazebrook, K., et al. 1996, ApJS, 107, 1, doi: 10.1086/192352

Bagley, M. B., Finkelstein, S. L., Koekemoer, A. M., et al. 2022, arXiv e-prints, arXiv:2211.02495. https://arxiv.org/abs/2211.02495

Barro, G., Faber, S. M., Koo, D. C., et al. 2017, ApJ, 840, 47, doi: 10.3847/1538-4357/aa6b05

Bournaud, F., Perret, V., Renaud, F., et al. 2014, ApJ, 780, 57, doi: 10.1088/0004-637X/780/1/57

Buitrago, F., Conselice, C. J., Epinat, B., et al. 2014, MNRAS, 439, 1494, doi: 10.1093/mnras/stu034

Buitrago, F., Trujillo, I., Conselice, C. J., & Häußler, B. 2013, MNRAS, 428, 1460, doi: 10.1093/mnras/sts124

Ceverino, D., Dekel, A., & Bournaud, F. 2010, MNRAS, 404, 2151, doi: 10.1111/j.1365-2966.2010.16433.x

Ceverino, D., Primack, J., & Dekel, A. 2015, MNRAS, 453, 408, doi: 10.1093/mnras/stv1603

Chen, T., Kornblith, S., Norouzi, M., & Hinton, G. 2020a, arXiv e-prints, arXiv:2002.05709. https://arxiv.org/abs/2002.05709

Chen, Z., Faber, S. M., Koo, D. C., et al. 2020b, ApJ, 897, 102, doi: 10.3847/1538-4357/ab9633

Conselice, C. J. 2003, ApJS, 147, 1, doi: 10.1086/375001

Costantin, L., Méndez-Abreu, J., Corsini, E. M., et al. 2020, ApJL, 889, L3, doi: 10.3847/2041-8213/ab6459

Costantin, L., Pérez-González, P. G., Méndez-Abreu, J., et al. 2021, ApJ, 913, 125, doi: 10.3847/1538-4357/abef72

—. 2022a, ApJ, 929, 121, doi: 10.3847/1538-4357/ac5a57

Costantin, L., Pérez-González, P. G., Vega-Ferrero, J., et al. 2022b, arXiv e-prints, arXiv:2208.00007. https://arxiv.org/abs/2208.00007

Dimauro, P., Daddi, E., Shankar, F., et al. 2022, MNRAS, 513, 256, doi: 10.1093/mnras/stac884

Ferreira, L., Adams, N., Conselice, C. J., et al. 2022a, ApJL, 938, L2, doi: 10.3847/2041-8213/ac947c

Ferreira, L., Conselice, C. J., Sazonova, E., et al. 2022b, arXiv e-prints, arXiv:2210.01110. https://arxiv.org/abs/2210.01110

Finkelstein, S. L., Dickinson, M., Ferguson, H. C., et al. 2017, The Cosmic Evolution Early Release Science (CEERS) Survey, JWST Proposal ID 1345. Cycle 0 Early Release Science

Finkelstein, S. L., Bagley, M. B., Ferguson, H. C., et al. 2022a, arXiv e-prints, arXiv:2211.05792, doi: 10.48550/arXiv.2211.05792

—. 2022b, arXiv e-prints, arXiv:2211.05792. https://arxiv.org/abs/2211.05792

Freundlich, J., Combes, F., Tacconi, L. J., et al. 2019, A&A, 622, A105, doi: 10.1051/0004-6361/201732223

Gardner, J. P., Mather, J. C., Clampin, M., et al. 2006, SSRv, 123, 485, doi: 10.1007/s11214-006-8315-7

Genzel, R., Tacconi, L. J., Gracia-Carpio, J., et al. 2010, MNRAS, 407, 2091, doi: 10.1111/j.1365-2966.2010.16969.x

Ginzburg, O., Huertas-Company, M., Dekel, A., et al. 2021, MNRAS, 501, 730, doi: 10.1093/mnras/staa3778

Grogin, N. A., Kocevski, D. D., Faber, S. M., et al. 2011a, ApJS, 197, 35, doi: 10.1088/0067-0049/197/2/35

—. 2011b, ApJS, 197, 35, doi: 10.1088/0067-0049/197/2/35

Guo, Y., Ferguson, H. C., Bell, E. F., et al. 2015, ApJ, 800, 39, doi: 10.1088/0004-637X/800/1/39

Guo, Y., Rafelski, M., Bell, E. F., et al. 2018, ApJ, 853, 108, doi: 10.3847/1538-4357/aaa018

Hayat, M. A., Stein, G., Harrington, P., Lukić, Z., & Mustafa, M. 2021, ApJL, 911, L33, doi: 10.3847/2041-8213/abf2c7

Hinton, G., Vinyals, O., & Dean, J. 2015, arXiv e-prints, arXiv:1503.02531. https://arxiv.org/abs/1503.02531

Huertas-Company, M., Pérez-González, P. G., Mei, S., et al. 2015, ApJ, 809, 95, doi: 10.1088/0004-637X/809/1/95

Huertas-Company, M., Rodriguez-Gomez, V., Nelson, D., et al. 2019, MNRAS, 489, 1859, doi: 10.1093/mnras/stz2191

Huertas-Company, M., Guo, Y., Ginzburg, O., et al. 2020, MNRAS, 499, 814, doi: 10.1093/mnras/staa2777

Kartaltepe, J. S., Rose, C., Vanderhoof, B. N., et al. 2022, arXiv e-prints, arXiv:2210.14713. https://arxiv.org/abs/2210.14713

Kassin, S. A., Weiner, B. J., Faber, S. M., et al. 2012, ApJ, 758, 106, doi: 10.1088/0004-637X/758/2/106

Kodra, D., Andrews, B. H., Newman, J. A., et al. 2022, arXiv e-prints, arXiv:2210.01140. https://arxiv.org/abs/2210.01140

Koekemoer, A. M., Faber, S. M., Ferguson, H. C., et al. 2011, ApJS, 197, 36, doi: 10.1088/0067-0049/197/2/36

Marinacci, F., Pakmor, R., & Springel, V. 2014, MNRAS, 437, 1750, doi: 10.1093/mnras/stt2003

McInnes, L., Healy, J., & Melville, J. 2018, arXiv e-prints, arXiv:1802.03426. https://arxiv.org/abs/1802.03426

Nelson, D., Springel, V., Pillepich, A., et al. 2019a, Computational Astrophysics and Cosmology, 6, 2, doi: 10.1186/s40668-019-0028-x

Nelson, D., Pillepich, A., Springel, V., et al. 2019b, MNRAS, 490, 3234, doi: 10.1093/mnras/stz2306

Peng, C. Y., Ho, L. C., Impey, C. D., & Rix, H.-W. 2010, AJ, 139, 2097, doi: 10.1088/0004-6256/139/6/2097

Pillepich, A., Nelson, D., Springel, V., et al. 2019, MNRAS, 490, 3196, doi: 10.1093/mnras/stz2338

Pozzetti, L., Bolzonella, M., Zucca, E., et al. 2010, A&A, 523, A13, doi: 10.1051/0004-6361/200913020

Robertson, B. E., Tacchella, S., Johnson, B. D., et al. 2023, ApJL, 942, L42, doi: 10.3847/2041-8213/aca086

Rodriguez-Gomez, V., Snyder, G. F., Lotz, J. M., et al. 2019, MNRAS, 483, 4140, doi: 10.1093/mnras/sty3345

Sarmiento, R., Huertas-Company, M., Knapen, J. H., et al. 2021, ApJ, 921, 177, doi: 10.3847/1538-4357/ac1dac

Scoville, N., Aussel, H., Brusa, M., et al. 2007, ApJS, 172, 1, doi: 10.1086/516585

Simons, R. C., Kassin, S. A., Weiner, B. J., et al. 2017, ApJ, 843, 46, doi: 10.3847/1538-4357/aa740c

Tomassetti, M., Dekel, A., Mandelker, N., et al. 2016, MNRAS, 458, 4477, doi: 10.1093/mnras/stw606

van der Wel, A., Franx, M., van Dokkum, P. G., et al. 2014a, ApJ, 788, 28, doi: 10.1088/0004-637X/788/1/28

van der Wel, A., Chang, Y.-Y., Bell, E. F., et al. 2014b, ApJL, 792, L6, doi: 10.1088/2041-8205/792/1/L6

Wisnioski, E., Förster Schreiber, N. M., Wuyts, S., et al. 2015, ApJ, 799, 209, doi: 10.1088/0004-637X/799/2/209

Wu, Z., Xiong, Y., Yu, S., & Lin, D. 2018, arXiv e-prints, arXiv:1805.01978. https://arxiv.org/abs/1805.01978

Zanisi, L., Huertas-Company, M., Lanusse, F., et al. 2021, MNRAS, 501, 4359, doi: 10.1093/mnras/staa3864

Zhang, H., Primack, J. R., Faber, S. M., et al. 2019, MNRAS, 484, 5170, doi: 10.1093/mnras/stz339

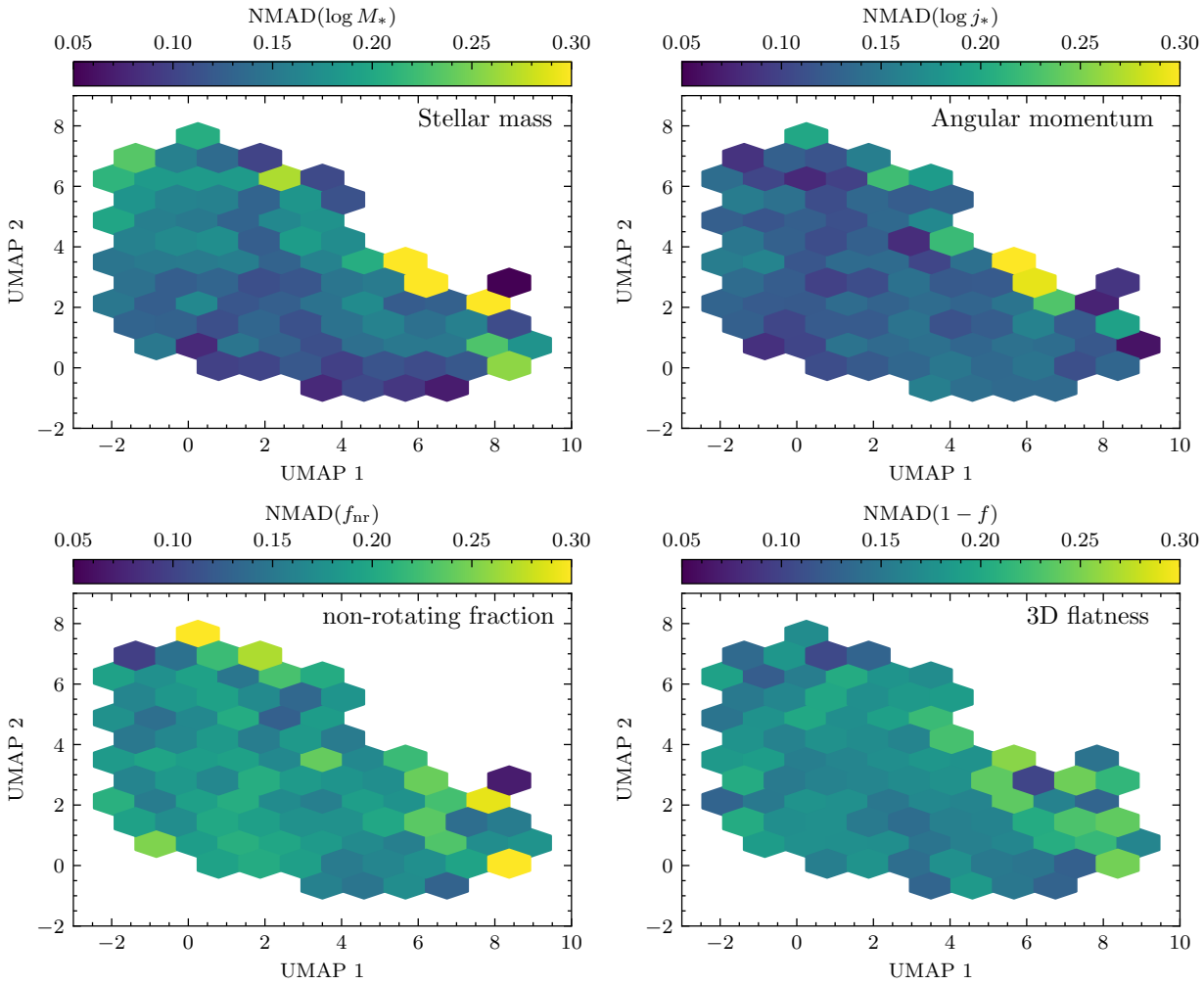

Figure A1. UMAP visualization for all the TNG50 galaxy images in our dataset color-coded by the distribution of several physical properties extracted from the TNG50 simulation. From left to right and top to bottom, the different panels show the UMAP plane color-coded by the NMAD of: the logarithm of the total stellar mass ($\log M_*[\mathrm{M}_\odot]$), the logarithm of the specific angular momentum of the stars ($\log j_*[\mathrm{kpc\ km\ s^{-1}}]$), the mass fraction in non-rotating stars ($f_{\rm nr}$) and the galaxy flatness ($1-f$).

APPENDIX

A. SCATTER OF PHYSICAL AND PHOTOMETRIC PARAMETERS IN THE UMAP VISUALIZATION

In Figure A1 and Figure A2, we show the variability in the physical and photometric parameters, respectively, in the UMAP visualization shown in Figure 9 in subsection 3.5. The scatter is quantified as a normalized median absolute deviation, denoted here as NMAD. The median absolute deviation (MAD), defined as $\mathrm{MAD(y)} = median(|y - median(y)|)$, is a robust measure of the variability of a univariate sample of quantitative data. The MAD is less affected by outliers and non-gaussianity than the typical variance and standard deviation. To facilitate the comparison between different variables, we normalized the MAD by the dynamical range of the data, defined as the percentile range containing 98% of the data. The resulting NMAD is an indicator of the variability of the data that, in this case, shows how informative the correlation with the different parameters shown in Figure 9 are. Values of NMAD $\lesssim 0.2$ are indicative of a low variability of the data.

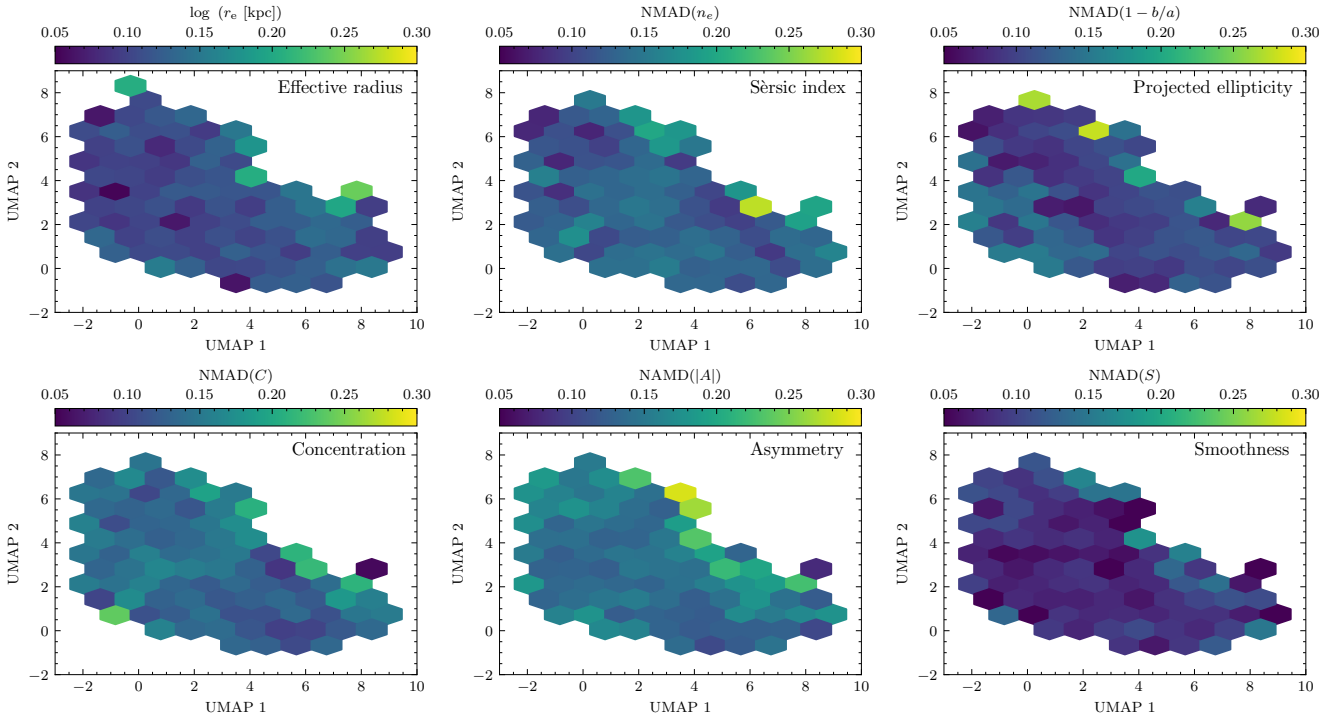

Figure A2. UMAP visualization for all the TNG50 galaxy images in our dataset color-coded by the distribution of several morphological and photometric parameters. From left to right and top to bottom, the different panels show the UMAP plane color-coded by the NMAD of: the logarithm of the effective radius (r_e[kpc]), the Sèrsic index (n_e), the ellipticity based on Sèrsic fit ($1 - b/a$), the concentration (C), the asymmetry (A) and the smoothness (S).

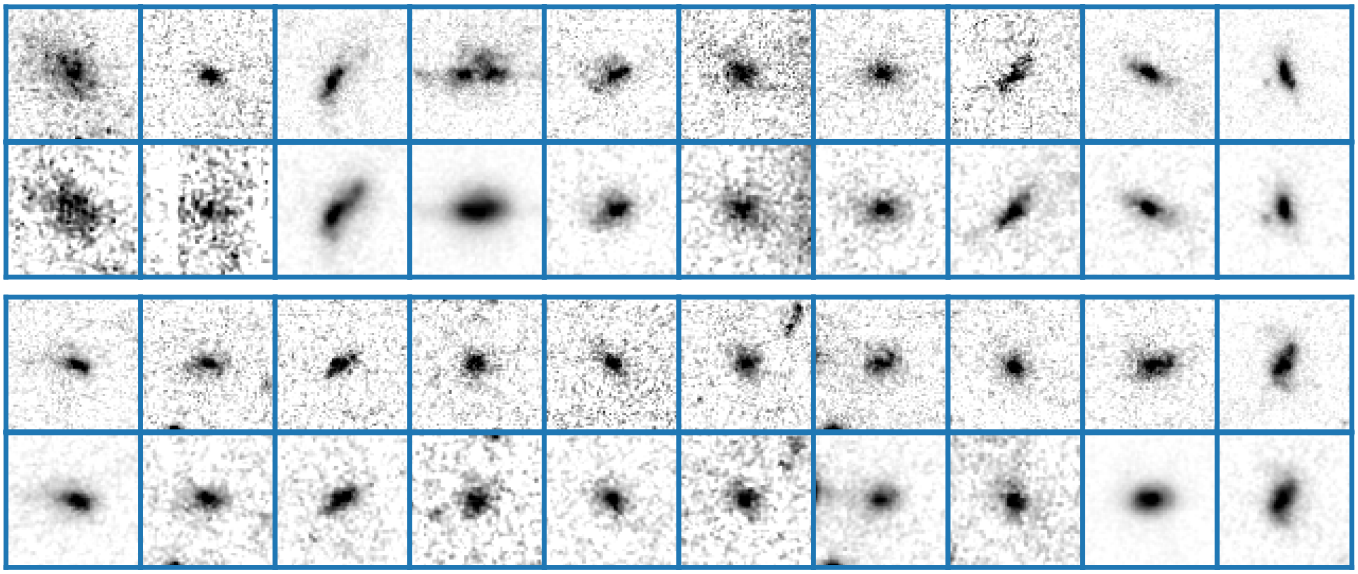

Figure B1. Examples of 20 galaxy images considered as *oblate disks* candidates in the VISUAL dataset. Images are framed in color according to the visual classification (blue for *Disk*). Each galaxy image is shown in the two F200W (top row) and the F356W (bottom row) filters. See subsubsection 5.2.2 for more details.

B. EXAMPLES OF OBSERVED JWST GALAXY IMAGES

In Figure B1, Figure B2, Figure B3 and Figure B4, we show some examples of galaxies in the CEERS and the VISUAL datasets according to different selection criteria.

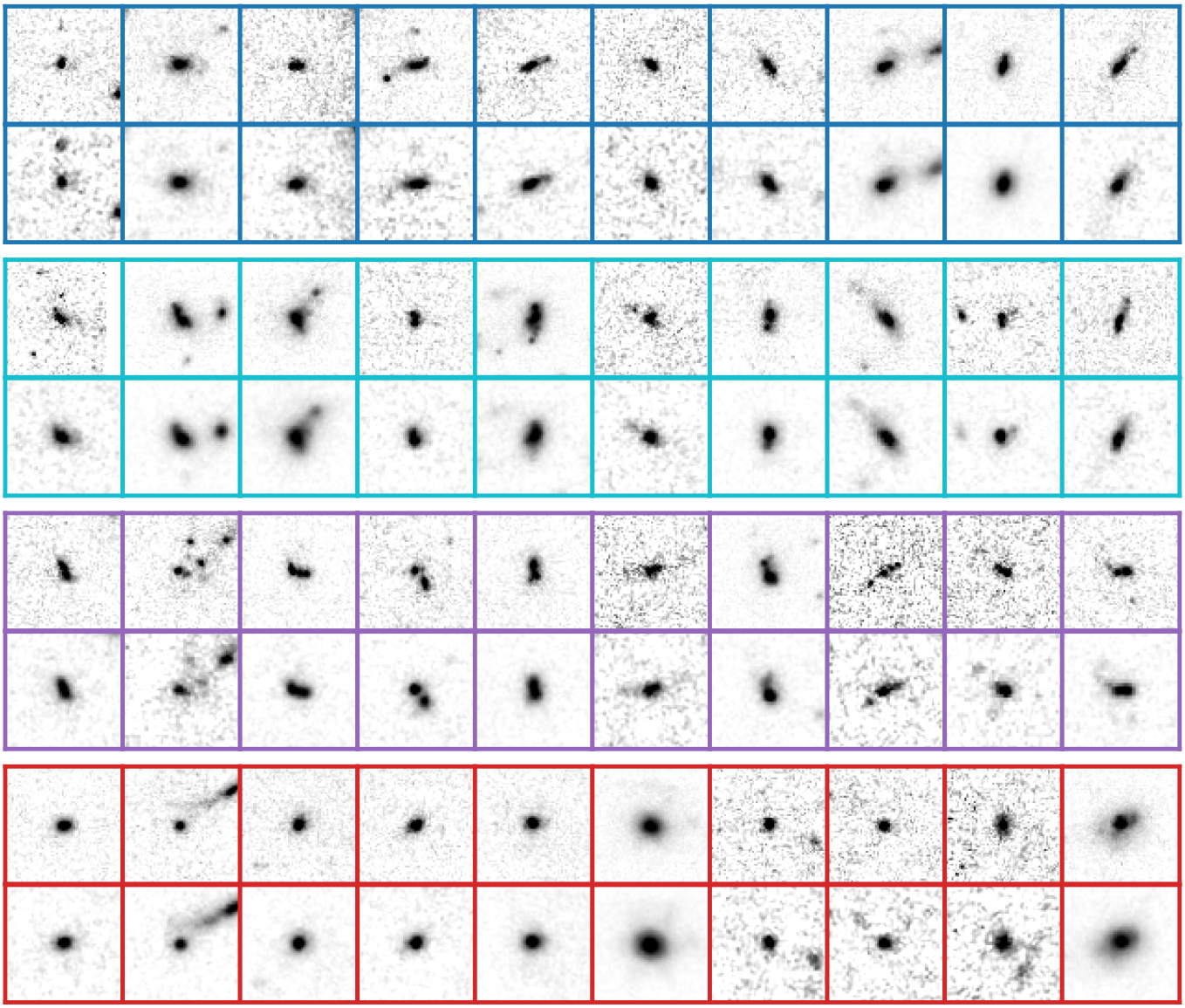

Figure B2. Examples of VISUAL galaxy images denoted as *non-oblate* galaxies. Different rows show examples of the various visual morphologies according to the VISUAL project, as follows: images framed in blue show examples of galaxies classified as *Disks*; images framed in cyan show examples of galaxies classified as *Disks+Irr*; images framed in purple show examples of galaxies classified as *Irr*; images framed in red show examples of *Sph* galaxies. For each galaxy type, we show images in the F200W (upper row) and F356W (bottom row) bands.

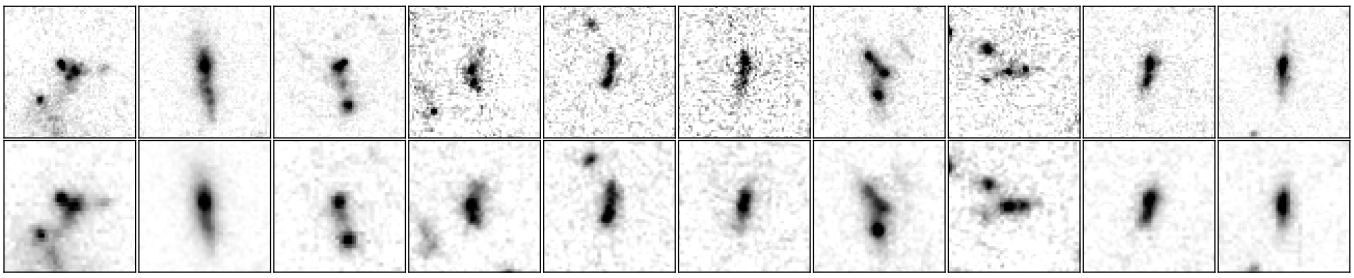

Figure B3. Examples of double nucleus or interacting galaxy images extracted from the CEERS dataset in the F200W (upper row) and F356W (bottom row) filters. Galaxies are selected in the UMAP plane as follows: $1 \lesssim$ UMAP $1 \lesssim 2$ and $0 \lesssim$ UMAP $2 \lesssim 1$. See Figure 10 and subsection 4.1 for more details.

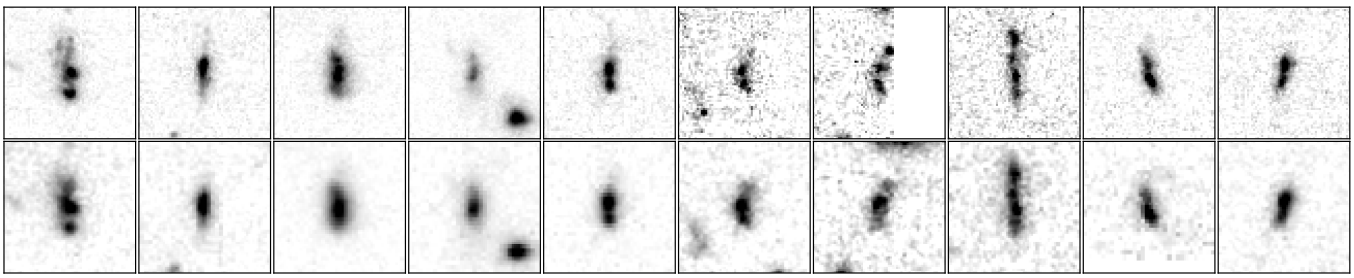

Figure B4. Examples of double nucleus or interacting galaxy images extracted from the VISUAL dataset in the F200W (upper row) and F356W (bottom row) filters. Galaxies are selected in the UMAP plane as follows: $1 \lesssim$ UMAP $1 \lesssim 2$ and $0 \lesssim$ UMAP $2 \lesssim 1$. See Figure 10, subsection 4.2 and subsubsection 5.2.4 for more details.